OE 6000 (REV. 9-66)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

(TOP)

ERIC ACCESSION NO. ED 013 343

CLEARINGHOUSE ACCESSION NUMBER	RESUME DATE	P.A.	T.A.
	22-01-68	48	

IS DOCUMENT COPYRIGHTED? YES ☒ NO ☐
ERIC REPRODUCTION RELEASE? YES ☒ NO ☐

TITLE
The Foreign Language Attainments of Language Majors in the Senior Year--A Survey Conducted in U.S. Colleges and Universities.

PERSONAL AUTHOR(S)
Carroll, John B.

INSTITUTION (SOURCE)
SOURCE CODE MGG30940

REPORT/SERIES NO. BR-5-1255

OTHER SOURCE
SOURCE CODE

OTHER REPORT NO.

OTHER SOURCE
SOURCE CODE

OTHER REPORT NO.

PUB'L. DATE - -67
CONTRACT/GRANT NUMBER OEC-4-14-048

PAGINATION, ETC.
EDRS PRICE MF-$1.25 HC-$10.92 273p.

RETRIEVAL TERMS
*MODERN LANGUAGES, *RESEARCH, *TEACHER EDUCATION, *LANGUAGE PROFICIENCY, *COLLEGE LANGUAGE PROGRAMS, PROGRAM EVALUATION, COLLEGE STUDENTS, LANGUAGE TESTS, LANGUAGE INSTRUCTION, LANGUAGE ABILITY, STUDENT CHARACTERISTICS, STUDENT MOTIVATION, STUDENT EVALUATION, MEASUREMENT, STUDY ABROAD, FLES PROGRAMS,

MODERN LANGUAGE APTITUDE TEST(MLAT), FOREIGN SERVICE INSTITUTE

IDENTIFIERS
(FSI), FSI PROFICIENCY RATINGS, MLA PROFICIENCY TESTS,

ABSTRACT
A study was made of the levels of proficiency attained by foreign language majors in U.S. colleges and universities, and of the factors associated with the attainments of these levels. The MLA Foreign Language Proficiency Tests for Teachers and Advanced Students were administered in 1965 to 2,775 seniors majoring in French, German, Italian, Russian, or Spanish at 203 institutions. Also used in the study were the Carroll-Sapon Modern Language Aptitude Test and questionnaires for majors and department chairmen. The resulting data provided new evidence on a number of issues significant in the selection and training of future language teachers, though conclusions must be drawn with caution. Among the students sampled there was a generally low achievement on listening and speaking skills. There was evidence that time spent abroad has a potent effect on a student's language skills, that students of Spanish or French who started the language in elementary school and continued it tended to have an advantage over other majors, that those from homes where foreign language was spoken attained greater competence, and that many low-aptitude students are able to compensate by diligent study and practice or because of special opportunities such as study abroad. Males and females were equal in language learning ability. Students at larger institutions outperformed those at smaller ones, and students at private

INSTRUCTIONS FOR COMPLETING ERIC REPORT RESUME

The resume is used to identify summary data and information about each document acquired, processed, and stored within the ERIC system. In addition to serving as a permanent record of the document in the collection, the resume is also a means of dissemination. All fields of the form must be completed in the allotted spaces, but inapplicable fields should be left blank. The following instructions are keyed to the line numbers appearing in the left margin of the form:

TOP LINE. ERIC Accession No. Leave blank. A permanent ED number will be assigned to each resume and its corresponding document as they are processed into the ERIC system.

LINE 001. Clearinghouse Accession No. For use only by ERIC Clearinghouses. Enter the alpha code and 6-digit document number.

Resume Date. In numeric form, enter month, day, and year that resume is completed. *(Example: 07 14 66)*

P.A. Leave blank.

T.A. Leave blank.

Copyright. Check appropriate block to denote presence of copyrighted material within the document.

ERIC Reproduction Release. Check appropriate block to indicate that ERIC has permission to reproduce the document and its resume form.

LINES 100-103. Title. Enter the complete document title, including subtitles if they add significant information. Where applicable, also enter volume number or part number, and the type of document *(Final Report, Interim Report, Thesis, etc.)*.

LINE 200. Personal Author(s). Enter personal author(s), last name first. *(Example: Doe, John J.)* If two authors are given, enter both. *(Example: Doe, John J. Smith, Ted)*. If there are three or more authors, list only one followed by "and others."

LINE 300. Institution *(Source)*. Enter the name of the organization which originated the report. Include the address *(city and State)* and the subordinate unit of the organization. *(Example: Harvard Univ., Cambridge, Mass., School of Education.)*

Source Code. Leave blank.

LINE 310. Report/Series No. Enter any unique number assigned to the document by the institutional source. *(Example: SC-1234)*

LINE 320. Other Source. Use only when a second source is associated with the document. Follow instructions for Line 300 above.

Source Code. Leave blank.

LINE 330. Other Report No. Enter document number assigned by the second source.

LINE 340. Other Source. Use only when a third source is associated with the document. Follow instructions for Line 300 above.

Source Code. Leave blank.

LINE 350. Other Report No. Enter document number assigned by the third source.

LINE 400. Publication Date. Enter the day, month, and year of the document. *(Example: 12 Jun 66)*

Contract/Grant Number. Applicable only for documents generated from research sponsored by the U.S. Office of Education. Enter appropriate contract or grant number and its prefix. *(Example: OEC-1-6-061234-0033)*

LINES 500-501. Pagination, etc. Enter the total number of pages of the document, including illustrations and appendixes. *(Example: 115p.)* USE THIS SPACE FOR ADDITIONAL INFORMATION PERTINENT TO THE DOCUMENT, such as publisher, journal citation, and other contract numbers.

LINES 600-606. Retrieval Terms. Enter the important subject terms *(descriptors)* which, taken as a group, adequately describe the contents of the document.

LINE 607. Identifiers. Enter any additional important terms, more specific than descriptors, such as trade names, equipment model names and numbers, organization and project names, discussed in the document.

LINES 800-822. Abstract. Enter an informative abstract of the document. Its style and content must be suitable for public announcement and dissemination.

U.S. GOVERNMENT PRINTING OFFICE : 1966 O—231-551

OE 6000 (REV. 9-66)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

(TOP)

ERIC ACCESSION NO. ED 013 343

CLEARINGHOUSE ACCESSION NUMBER

RESUME DATE — —

P.A.

T.A.

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☐

ERIC REPRODUCTION RELEASE? YES ☐ NO ☐

001

TITLE

100
101
102
103

PERSONAL AUTHOR(S)

200

INSTITUTION (SOURCE)

SOURCE CODE

300

310 REPORT/SERIES NO.

OTHER SOURCE

SOURCE CODE

320

330 OTHER REPORT NO.

OTHER SOURCE

SOURCE CODE

340

350 OTHER REPORT NO.

400 PUB'L. DATE — —

CONTRACT/GRANT NUMBER

PAGINATION, ETC.

500
501

RETRIEVAL TERMS

600
601
602
603
604
605
606

IDENTIFIERS

607

ABSTRACT

800 822 institutions did better than those at public ones. Statistical data
801 823 are presented in 99 tables and 13 figures and other background
802 824 information is contained in 5 appendixes. (AM)

803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822

OE 6000 (Rev. 9-66)

INSTRUCTIONS FOR COMPLETING ERIC REPORT RESUME

The resume is used to identify summary data and information about each document acquired, processed, and stored within the ERIC system. In addition to serving as a permanent record of the document in the collection, the resume is also a means of dissemination. All fields of the form must be completed in the allotted spaces, but inapplicable fields should be left blank. The following instructions are keyed to the line numbers appearing in the left margin of the form:

TOP LINE. ERIC Accession No. Leave blank. A permanent ED number will be assigned to each resume and its corresponding document as they are processed into the ERIC system.

LINE 001. Clearinghouse Accession No. For use only by ERIC Clearinghouses. Enter the alpha code and 6-digit document number.

Resume Date. In numeric form, enter month, day, and year that resume is completed. *(Example: 07 14 66)*

P.A. Leave blank.

T.A. Leave blank.

Copyright. Check appropriate block to denote presence of copyrighted material within the document.

ERIC Reproduction Release. Check appropriate block to indicate that ERIC has permission to reproduce the document and its resume form.

LINES 100–103. Title. Enter the complete document title, including subtitles if they add significant information. Where applicable, also enter volume number or part number, and the type of document *(Final Report, Interim Report, Thesis, etc.)*.

LINE 200. Personal Author(s). Enter personal author(s), last name first. *(Example: Doe, John J.)* If two authors are given, enter both. *(Example: Doe, John J. Smith, Ted)*. If there are three or more authors, list only one followed by "and others."

LINE 300. Institution *(Source)*. Enter the name of the organization which originated the report. Include the address *(city and State)* and the subordinate unit of the organization. *(Example: Harvard Univ., Cambridge, Mass., School of Education.)*

Source Code. Leave blank.

LINE 310. Report/Series No. Enter any unique number assigned to the document by the institutional source. *(Example: SC-1234)*

LINE 320. Other Source. Use only when a second source is associated with the document. Follow instructions for Line 300 above.

Source Code. Leave blank.

LINE 330. Other Report No. Enter document number assigned by the second source.

LINE 340. Other Source. Use only when a third source is associated with the document. Follow instructions for Line 300 above.

Source Code. Leave blank.

LINE 350. Other Report No. Enter document number assigned by the third source.

LINE 400. Publication Date. Enter the day, month, and year of the document. *(Example: 12 Jun 66)*

Contract/Grant Number. Applicable only for documents generated from research sponsored by the U.S. Office of Education. Enter appropriate contract or grant number and its prefix. *(Example: OEC-1-6-061234-0033)*

LINES 500–501. Pagination, etc. Enter the total number of pages of the document, including illustrations and appendixes. *(Example: 115p.)* USE THIS SPACE FOR ADDITIONAL INFORMATION PERTINENT TO THE DOCUMENT, such as publisher, journal citation, and other contract numbers.

LINES 600–606. Retrieval Terms. Enter the important subject terms *(descriptors)* which, taken as a group, adequately describe the contents of the document.

LINE 607. Identifiers. Enter any additional important terms, more specific than descriptors, such as trade names, equipment model names and numbers, organization and project names, discussed in the document.

LINES 800–822. Abstract. Enter an informative abstract of the document. Its style and content must be suitable for public announcement and dissemination.

U.S. GOVERNMENT PRINTING OFFICE : 1966 O—231-551

THE FOREIGN LANGUAGE ATTAINMENTS OF LANGUAGE MAJORS IN THE SENIOR YEAR:

A SURVEY CONDUCTED IN U. S. COLLEGES AND UNIVERSITIES

FINAL REPORT
for
Contract OE-4-14-048
with the U. S. Office
of Education,
Department of Health,
Education and Welfare

PRINCIPAL INVESTIGATOR
John B. Carroll
Professor of Educational
Psychology

STAFF
John L. D. Clark
Thomas M. Edwards
Fannie A. Handrick

Laboratory for Research in Instruction
Graduate School of Education
Harvard University

Cambridge, Massachusetts 02138
1967

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

THE FOREIGN LANGUAGE ATTAINMENTS OF LANGUAGE MAJORS IN THE SENIOR YEAR:

A SURVEY CONDUCTED IN U. S. COLLEGES AND UNIVERSITIES

FINAL REPORT
for
Contract OE-4-14-048
with the U. S. Office
of Education,
Department of Health,
Education and Welfare

PRINCIPAL INVESTIGATOR
John B. Carroll
Professor of Educational
Psychology

STAFF
John L. D. Clark
Thomas M. Edwards
Fannie A. Handrick

This study was supported in part by funds provided by the U. S. Office of Education under Title VI, Section 602, Public Law 85-864, Contract No. OE-4-14-048. Reproduction in whole or in part permitted for any purpose of the United States Government.

Laboratory for Research in Instruction
Graduate School of Education
Harvard University

Cambridge, Massachusetts 02138
1967

COPYRIGHT 1967 BY THE
PRESIDENT AND FELLOWS OF HARVARD COLLEGE

PRINTED BY THE HARVARD UNIVERSITY PRINTING OFFICE
Job No. K 357

TABLE OF CONTENTS

Page
LIST OF TABLES vi

LIST OF FIGURES ix

PREFACE xiv

Chapter
I. INTRODUCTION TO THE STUDY 1

1. Purposes 1
2. Overview of the Study 3

II. MEASUREMENTS AND PROCEDURES 5

1. Instruments 5
2. The Establishment of Tentative MLA Proficiency Test Equivalences with FSI (Foreign Service Institute) Absolute Proficiency Ratings 10
3. The Design of the Sample 15
4. Procedures in Data Collection 27

III. OUTCOMES OF THE SAMPLING DESIGN 31

1. Institutional Response Rates 31
2. Student Response Rates 34
3. Completeness of Data and Preparation of the Sample 37

IV. RESULTS FROM TESTS OF FOREIGN LANGUAGE ATTAINMENT: THE TESTED SAMPLE 40

1. Distributional Statistics, All Cases 40
2. Distributional Statistics, "Regular" and "Odd" Type Cases 46
3. Norms for Regular Cases 49
4. Intertest Correlations 49

V. ESTIMATES OF FOREIGN LANGUAGE SKILL ATTAINMENTS IN THE TOTAL POPULATION 59

1. Bias due to Student Non-Response 59
2. Variations in Foreign Language Skill Associated with Stratification Variables in the Sample Design 66
3. Variations in Foreign Language Skill Associated with Geographical Region 77
4. Estimated National Norms for MLA Skills Tests 79

VI. STUDENT FACTORS IN FOREIGN LANGUAGE SKILL ATTAINMENT 101

1. Introduction 101
2. Sex and Age 101
3. Foreign Language Aptitude 103
4. Plans, Interests, and Motives in Foreign Language Study 109
5. Major Dimensions of a "Typology" of Foreign Language Majors 122

VII. EXPERIENCE AND TRAINING FACTORS IN FOREIGN LANGUAGE SKILL ATTAINMENT 126

1. Introduction 126
2. Time of Beginning Foreign Language Study 126
3. Instructional Variables 131
4. Time Spent Abroad 137

Table of Contents (continued)

Page
5. Use of the Major Foreign Language at Home 141
6. Other Experiences in the Foreign Language 141
7. Other Foreign Languages Taken Besides Major Language 145

VIII. REGRESSION ANALYSES FOR PREDICTING MLA SKILLS TEST SCORES FROM COMPOSITES OF VARIABLES 151

1. Introduction 151
2. Multiple Regression Prediction of MLA Skills Test Scores from MLAT and Three Background Variables, French Majors 152
3. Multiple Regression Techniques Applied to Prediction of Each MLA Skills Test Variable, in Each Language 155
4. Canonical Regression Analysis of MLA Skills Test Scores with MLAT and Three Background Variables as Predictors, French Majors 162
5. Canonical Regression Techniques Applied to Prediction of Linear Composites of MLA Skills Test Scores 163
6. The Stability of Regression Systems over Institution Size Strata and Institution Types 171

IX. CHARACTERISTICS OF MODERN FOREIGN LANGUAGE DEPARTMENTS IN RELATION TO STUDENT ACHIEVEMENT 173

1. Introduction 173
2. Dimensions of Instruction in Subdepartments 173
3. Departmental Characteristics and Student Achievement 177

X. RESULTS FROM THE MLA TEACHER PREPARATION TESTS 181

1. The Sample 181
2. Estimated National Norms for Teacher Preparation Tests 184
3. Intercorrelations of MLA Teacher Preparation Tests and Correlations with MLA Skills Tests 184
4. Variables Associated with Performance on the MLA Teacher Preparation Tests 190

XI. SUMMARY AND RECOMMENDATIONS 197

1. Purpose and Design of the Study 197
2. Outcomes of the Sampling Design 198
3. Levels of Proficiency Attained by Students in the Tested Sample 199
4. Estimates of Levels of Foreign Language Proficiency Attained in the National Population 199
5. Student Factors in Foreign Language Skill Attainment 201
6. Experience and Training Factors in FL Skill Attainment 202
7. Comparisons of Influences of Student and Experiential Variables 203
8. Foreign Language Department Characteristics in Relation to Student Achievement 204
9. Results from the MLA Teacher Preparation Tests Administered to Students Enrolled in Teacher Preparation Programs 204
10. Recommendations for the Improvement of the Education of Foreign Language Majors 205
11. Recommendations for Further Studies of Foreign Language Education 206

REFERENCES 209

Table of Contents (continued)

APPENDICES

Page

A. An Analysis of Data from the New York Pilot Study 213

B. Questionnaire for Foreign Language Majors 224

C. Questionnaire for Chairmen of Foreign Language Departments 228

D. Administrative Materials Mailed to Participating Institutions 234

1. Memorandum to Presidents of Institutions of Higher Learning 234
2. Memorandum Describing the Project 235
3. Institution Information Form 237
4. Instructions and Suggestions to Coordinator for Administering Test Materials under the College Foreign Language Testing Project, Spring 1965 239
5. Order Form for Administration of MLA Foreign Language Proficiency Tests under the College Foreign Language Testing Project 244
6. A Letter to College Seniors Graduating in the Spring of 1965 with Degrees in French, Italian, Spanish, German, or Russian 246
7. Directions for Administering the Questionnaire for Foreign Language Majors 250

E. Score Reports 251

1. Memorandum to the Foreign Language Majors who Took the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students in the Spring of 1965, dated October 5, 1965 251
2. Letter to Department Chairmen, dated October 12, 1965 252
3. Interpretative Leaflet, MLA FL Proficiency Tests 253

LIST OF TABLES

Page

Table

2.1 Cases Tested by FSI and Ranges of FSI Ratings 12

2.2 Basic Data for Equivalency of FSI "Absolute Proficiency Ratings" and Scores on MLA Proficiency Tests for Teachers and Advanced Students 14

2.3 MLA Proficiency Test Equivalents (Y) and Parameters of Linear Equations Relating Them to FSI Ratings (X), with Ranges of Myers and Melton Ratings 16

2.4 1962-63 Population of U. S. Institutions Offering A.B. Degree Programs in Test Languages, with Numbers of Earned Degrees 19

2.5 Worksheet for First Sampling Allocation 21

2.6 Worksheet for First Sampling Allocation Adjusted for Projected Growth 23

2.7 Initial Sampling Design, and Institutional Response as of January 14, 1965 to the First Mailing of December 10, 1964 24

2.8 Worksheet for Second Sampling Allocation 25

2.9 Final Sampling Design (Numbers of Institutions) 26

3.1 Institutional Response Rates 32

3.2 Rate of Return of Favorable Replies from Institutions 33

3.3 Student Response Rates, by Strata, Institution Type, and Sex 35

3.4 Student Response Rates, by Language 38

4.1 Data from MLA Skills Test Score Distributions (All Cases Complete on 4 MLA Skills Tests) 41

4.2 Distribution Statistics for MLA Skills Test Scores for "Regular" and "Odd" Type Cases 47

4.3 Norms for MLA Advanced FL Proficiency Tests Based on "Regular" Cases Tested, 1965 College Senior Major Samples 50

4.4 Intercorrelations of MLA Skills Tests for CFLT National Sample, as Compared to Results for NDEA Institute Standardization Samples 52

4.5 Proportions of Common Factor Variance Accounted for by the MLA Skills Tests (Derived from Correlations for the CFLT Sample in Table 4.4) 53

4.6 Estimated Reliability Coefficients for MLA Tests Applied to CFLT Sample and Estimated Correlations Corrected for Attenuation (Derived from Data of Table 4.4) 54

4.7 Instances of Significant Departure from Linearity in Intercorrelations of MLA Skills Tests 58

List of Tables (continued)

Table — Page

5.1 Comparison of Foreign Language GPA's of Tested and Non-Tested Students at Fifteen Selected Institutions, in Order of Percentage Tested 61

5.2 Correlations Among MLA Skills Test Scores and Between MLA Skills Test Scores and Standardized Foreign Language Grade-Point-Averages, by Language, for Cases with Complete Test Scores 64

5.3 Study of Bias Due to Student Non-Response, and Estimates of Correlations for National Sample Data, for French, German, and Spanish Students 65

5.4 Mean Skills Test Scores, by Stratum and Institution Type 67

5.5 F-Ratios for Stratum and Institution-Type Effects for Selected Variables 71

5.6 Means of Selected Students and Instructional Variables, by Stratum and Institution Type 73

5.7 F-Ratios for Stratum and Institution-Type Effects for MLA Skills Test Scores, Without and With Adjustment for Variation in the Time the Student Began FL Study and in Amount of Travel and Study Abroad 78

5.8 Mean MLA Skills Test Scores by Geographical Region (Regular Cases) 80

5.9 Ratios Used in Estimation of National Norms
A. Ratios B/C, by language, test, stratum, and institution type
B. Ratios D/E, by stratum and institution type 82

5.10 Means and Standard Deviations for Estimated National Norms 88

5.11 Estimated National Norms for MLA Proficiency Test Skill Scores, College Senior FL Majors (Percentile Ranks of Converted Scores) 89

5.12 Estimated College Senior Major National Norms for Selected Percentile Ranks 93

5.13 Comparison of Estimated College Senior FL Major Norms with Pretest and Posttest NDEA Norms, in Terms of Means ($\bar{X}$) and Standard Deviations (σ) 97

5.14 Estimated Percentages of CFLT Population Attaining or Exceeding Each FSI Rating 99

5.15 Estimated Numbers and Percentages of CFLT Population Falling in Each of the Myers-Melton Qualification Categories 100

6.1 MLA Skills Scores by Sex 102

6.2 Age Distributions of FL Majors at Graduation, by Language 104

6.3 Mean MLA Skill Test Scores by Age Groups 105

List of Tables (continued)

Table — Page

6.4 Modern Language Aptitude Test scores, by Language, as compared with norms 106

6.5 Intercorrelations of MLAT Part Scores and Correlations with MLA Skill Scores 108

6.6 Chief Reasons Given for Majoring in a Foreign Language, By Sex, in Percentages 110

6.7 Levels at Which Majors Plan to Teach, By Sex, in Terms of Percentage 111

6.8 Specific Employment Prospects of Those Planning to Use FL Skills in Work other than Teaching, By Sex, in Terms of Percentages 113

6.9 Mean Importance Ratings Given to Foreign Language Skills by Students 114

6.10 Intercorrelations of FL Skill Importance Ratings and Correlations with MLA Skill Scores, By Language 116

6.11 First Two Canonical Correlations between FL Skill Importance Ratings and MLA Skill Scores, and Canonical Regression Weights 117

6.12 Means and S.D.'s of MLA Skills Test Scores by Vocational Plans (Sexes Combined), and by Major Language 119

6.13 Means and S.D.'s of MLA Skills Test Scores, by Reasons Other than Vocational for Choosing a Foreign Language Major, and by Major Language 121

6.14 Factor Loadings for Selected Questionnaire Responses (Orthogonal Rotated Axes) 123

7.1 Numbers and Percentages of "Regular" Cases, by Time of Beginning FL Study, by Language (Excluding Italian) (All cases complete on Four MLA Skills Tests) 127

7.2 Means and S.D.'s of MLA Skills Test Scores for Groups Starting FL Study at Different Educational Levels (All "Regular" Cases Complete on MLA Skills Test Scores) 128

7.3 Mean MLAT Scores for Groups Starting their Major Foreign Language at Different Times 130

7.4 Mean Semesters or Course Hours Taken by Groups Starting at Different Times, with Mean MLA Skills Test Scores 132

7.5 Regression System for Predicting MLA Skills Test Scores from Amount of Study at Different Educational Levels--FRENCH 133

7.6 Regression System for Predicting MLA Skills Test Scores from Amount of Study at Different Educational Levels--SPANISH 134

7.7 Data for Certain Instructional Variables and Correlations with MLA Skills Tests 136

List of Tables (continued)

Table — Page

7.8 Mean MLA Skills Test Scores for Groups with Different Amounts of Experience Abroad, by Language and by Sex 138

7.9 Mean Listening Test Scores (and N's) by Time of Beginning FL Study and by Time Spent Abroad 140

7.10 Mean MLA Skills Test Scores by Use of the FL at Home ("Regular" Cases Only) 142

7.11 Amount of Use of the FL in Jobs, with Friends, Etc., by Language and by Sex 143

7.12 Amount of Independent Reading, by Language and by Sex, with Mean Reading Test Scores for Language Groups 144

7.13 Other Languages Taken, By Time of Beginning Major Language (Excluding Latin and Ancient Greek) 146

7.14 Mean MLA Skills Test Scores for Groups with Study of Other Languages Besides their Major Languages 147

7.15 Effects Associated with the Study of Latin and/or Ancient Greek 149

8.1 Multiple Regression Analyses for Predicting MLA Skills Test Scores from MLAT Scores and Three Background Variables, French Majors, "Regular" Cases (N = 1039) 153

8.2 Multiple Regression Analyses for the MLA Listening Test (Beta Weights and Multiple R's) 157

8.3 Multiple Regression Analyses for the MLA Speaking Test (Beta Weights and Multiple R's) 158

8.4 Multiple Regression Analyses for the MLA Reading Test (Beta Weights and Multiple R's) 160

8.5 Multiple Regression Analyses for the MLA Writing Test (Beta Weights and Multiple R's) 161

8.6 Canonical Regression Analysis for Data of Table 8.1 (MLAT Subtests and Three Background Variables as Predictors, Four MLA Skills Tests as Criteria), French Majors (N = 1039) 163

8.7 Canonical Regression Results for the General Foreign Language Competence Factor 165

8.8 Canonical Regression Results for the WR (Superiority in Writing and Reading) Factor 166

8.9 Canonical Regression Results for the SW (Superiority in Productive Skills) Factor 167

8.10 Canonical Regression Results for the RS (Superiority in Reading and Speaking) Factor 168

8.11 First Canonical Regression Weights, by Stratum and Institution Types, for 14 Selected Predictor Variables and MLA Skills Tests, French "Regular" Cases 172

List of Tables (continued)

Table | Page
9.1 National Study: Classification of Sub-Departments by Type of Department and Type and Size of Institution 174

9.2 National Study: Correlations Between Four Factors of Audiolinguality and Mean Student Ratings of Classroom Practices 175

9.3 National Study: Advanced Instruction Variables 176

9.4 Numbers of Students Tested in Each of Two Groups by Language 177

9.5 Multiple Regression Data for Prediction of Mean Student Achievement from Selected Departmental Variables, N = 70 Spanish Subdepartments, Variables derived from Students who Started Study of Spanish in College (from Handrick, 1967) 179

10.1 Summary Statistics for MLA Teacher Preparation Tests, Norm Groups (From Table 4. 3), compared with Data from NDEA Institute Standardization Groups 182

10.2 Comparisons on MLA Skills Tests of Those Who Did and Did Not Take Any Teacher Preparation Tests 183

10.3 Estimated National Norms for MLA Teacher Preparation Tests, College Senior FL Majors (Percentile Ranks of Converted Scores) 185

10.4 Estimated College Senior Major National Norms, MLA Teacher Preparation Tests Converted Scores, for Selected Percentile Ranks 187

10.5 Intercorrelations of MLA Teacher Preparation Tests and Correlations with MLA Skills Tests, for CFLT Samples as Compared with Results from NDEA Institute Standardization Samples 188

10.6 Correlations of MLAT Subtests with MLA Teacher Preparation Tests 191

10.7 Means and Standard Deviations of MLA Teacher Preparation Test Scores, by Intentions to Teach at Various Levels, for French and Spanish Groups 192

10.8 Mean MLA Teacher Preparation Test Scores, by Language and by Time of Beginning Study of the Major Language 193

10.9 Multiple and Canonical Regression Analyses for Teacher Preparation Tests and Selected Predictor Variables 195

LIST OF FIGURES

Page

Figure

4.1 Frequency Distribution of Scores on MLA Skills Tests, French, Tested Sample 42

4.2 Frequency Distribution of Scores on MLA Skills Tests, German, Tested Sample 43

4.3 Frequency Distribution of Scores on MLA Skills Tests, Russian, Tested Sample 44

4.4 Frequency Distribution of Scores on MLA Skills Tests, Spanish, Tested Sample 45

4.5 Regression of French Listening Score on French Speaking Score 56

4.6 Regression of French Listening Score on French Writing Score 56

4.7 Regression of German Listening or Writing Score on German Speaking Score 56

4.8 Regression of Russian Speaking Score on Russian Listening Score 57

4.9 Regression of Spanish Reading Score on Spanish Listening Score 57

5.1 Estimated National Norms for CFLT Population, with MLA-FSI and Myers-Melton Equivalents, French 84

5.2 Estimated National Norms for CFLT Population, with MLA-FSI and Myers-Melton Equivalents, German 85

5.3 Estimated National Norms for CFLT Population, with MLA-FSI and Myers-Melton Equivalents, Russian 86

5.4 Estimated National Norms for CFLT Population, with MLA-FSI and Myers-Melton Equivalents, Spanish 87

PREFACE

This study had its origin in informal discussions that occurred in late 1963 between the writer and several far-sighted and research-minded individuals in the U. S. Office of Education--all professionally concerned with the improvement of foreign language teaching --who suggested that there was a need for a thoroughgoing survey of the preparation of foreign language teachers in the colleges and universities of the United States. In particular, I have in mind Drs. D. Lee Hamilton, Kenneth Mildenberger, and A. Bruce Gaarder, who pointed out that such a survey had recently become much more feasible due to the availability of the then new MLA Foreign Language Proficiency Tests for Teachers and Advanced Students, designed to measure foreign language skills and teacher competences at an advanced level. The idea of conducting such a survey had immediate appeal for me, particularly if it could be done on a large enough scale and with the collection of sufficient auxiliary data to enable one to assess in depth the factors associated with the attainment of foreign language proficiency. Such a survey seemed also made to order to contribute toward the carrying out of one of the provisions of the National Defense Education Act of 1958: "The Commissioner is authorized, directly or by contract, to make studies and surveys to determine the need for increased or improved instruction in modern foreign languages..." (Section 602). As such, it might provide a model for similar "national assessment" studies at other educational levels, or indeed in other curriculum areas besides foreign languages.

I therefore welcomed the opportunity to plan and conduct the study reported here. The gentlemen named above were of great assistance in helping me make arrangements for negotiating a contract with the Department of Health, Education and Welfare. A succession of coordinators in the Language Research Section of the Bureau of Research of the U. S. Office of Education--Drs. James Alatis, Augustus Koski, Irving R. Wershow, and Al Storm and his assistant Mr. William Higgins--have been exceedingly helpful in smoothing the way when it came to such matters as clearance of forms with the Bureau of the Budget, the obtaining of certain statistical data, and the negotiation of contract extensions that became necessary. To all of them I express deep gratitude.

At an early stage of the study, several people in the New York State Education Department helped in facilitating the participation of colleges and universities in that state in the pilot investigation conducted in the spring of 1964. Mr. Norman D. Kurland, then Consultant on College Proficiency Examinations, deserves special mention.

The Modern Language Association, with its concern for a host of matters connected with foreign language teaching, has given generous moral support to this investigation at all stages. Most of the individuals already named as being in the U. S. Office of Education were in fact members of the MLA who, from a professional standpoint, were acting in behalf of some of the general purposes of the MLA Foreign Language Program in promoting this study. After returning to the MLA central office from his post in the U. S. Office of Education, Dr. Mildenberger continued to evince great interest in the study and has been of much assistance in publicizing the preliminary results. Dr. John Hurt Fisher, as Secretary of the Association, has watched the progress of the study with a benign eye. Dr. Joseph G. Astman, and later, Mr. F. André Paquette and his assistant Mrs. Suzanne Tollinger, charged with various matters concerning the development and use of the MLA Foreign Language Proficiency Tests, have given much helpful information and advice.

It would have been virtually impossible to conduct the study without the cooperation of Educational Testing Service. Dr. Robert Solomon, Dr. Peter Loret, Miriam Bryan, and Olive Bray helped arrange what seemed to the writer the mammoth task of administering and scoring tests given on tight schedules at several hundred different institutions throughout the United States. I am very grateful to them.

Thanks also are due to Dr. James Frith and Miss Claudia Wilds of the Foreign Service Institute, Department of State, Washington, D. C., who managed to arrange, despite many obstacles, the giving of FSI language proficiency interviews to more than a hundred persons at seven different institutions in the eastern part of the country.

On June 20-21, 1966, a number of consultants were called together at Harvard to examine the preliminary results of the survey and give suggestions as to further questions that might be answered by suitable analyses of the data. I am thankful to all the members of that most interesting and fruitful conference. They were:

Dr. Edward Allen, College of Education, Ohio State University
Dr. Joseph G. Astman, Department of Foreign Languages and Literatures, Hofstra University
Dr. Dwight Bolinger, Department of Romance Languages, Harvard University
Dr. Nelson Brooks, Modern Language Department, Yale University
Dr. Alfred Hayes, Center for Applied Linguistics, Washington, D. C.
Dr. Wallace E. Lambert, Department of Psychology, McGill University
Dr. Peter Loret, Educational Testing Service
Dr. Andrea McHenry Mildenberger, Modern Language Association
Dr. Kenneth Mildenberger, Modern Language Association
Mr. F. André Paquette, Modern Language Association
Dr. Jack Stein, Department of Germanic Languages, Harvard University
Mrs. Suzanne Tollinger, Modern Language Association

At Harvard, I have received welcome support from Dean Theodore R. Sizer of the Graduate School of Education, Deans Franklin Ford and John Monro of Harvard College, and Dean Kathleen Elliott of Radcliffe College. Dr. Dean Whitla of the Office of Tests assisted in arranging testing at Harvard. Dr. Theodore Colton of the Harvard Medical School was especially helpful and generous at all stages in giving advice concerning the sampling design of the study; later, Professor William G. Cochran of the Department of Statistics helped to resolve some difficult problems of data analysis and interpretation. The facilities of the Harvard Computing Center were of enormous aid, as well as the programs developed by Drs. William Cooley, Paul Lohnes, and Kenneth J. Jones, all former students or faculty members of the Harvard Graduate School of Education.

Throughout the three-year period of the study, I have been fortunate in having the services of an excellent staff. As Research Assistant, Mr. (now Dr.) John L. D. Clark was primarily responsible for working out the procedural details of data collection for the study in both its first and second phases, also some of the statistical analysis. Mr. Thomas M. Edwards, Research Assistant, was primarily responsible for managing the details of statistical data analysis in the second, major phase of the study. Mrs. (now Dr.) Fannie A. Handrick has at various times served as a Research Assistant especially concerned with the analysis of data from the department chairmen's questionnaires; Chapter IX of this report is based on her doctoral dissertation. Mrs. Mary Klaaren served as a devoted Project Secretary throughout nearly the whole period of the study and performed many seemingly thankless clerical tasks in reducing the mountains of data to manageable proportions. Miss Marjorie M. Morse, Jr., Mrs. Ellen W. Davis, Mrs. Nancy Kinney Soderberg, and Mr. Robert Stryker gave expert attention through countless hours of coding, punching, card sorting, and running of statistical tabulations. During the final stages of the study, Miss Sally Chapin cheerfully allowed herself to be pressed into service to type the camera-ready copy of the report with its endless statistical tables. My regular secretary and administrative assistant, Mrs. Edith A. Alexander, was helpful in numerous ways. To all these people I am grateful in more ways than words can say.

Although I was assisted by a great many people, none of them should be held responsible for any errors of planning, execution, or interpretation that may be noted in this report. I accept sole responsibility for this study and the present report.

The report had to be written, unfortunately, at a time when I was in transition between Harvard University and a new post as Senior Research Psychologist at Educational Testing Service, Princeton, which I took up in February of this year. I am thankful to my colleagues at Harvard for continuing to provide me with the facilities and space to complete the project, and to my colleagues at Educational Testing Service for allowing me time away from other duties. But I am even more appreciative to my wife, Mrs. Mary S. Carroll, for allowing me time off from my responsibilities to her and our home in this rather difficult time of personal upheaval. To her I hope I can promise that writing pressures will never again be as heavy, but she has every right not to believe my promises.

John B. Carroll

Cambridge, Massachusetts, July 31, 1967

Special note: The original data of this investigation, save for information that would identify institutions and individuals tested, will be archived at the ERIC Clearinghouse on Foreign Languages, Modern Language Association, 62 Fifth Avenue, New York, N. Y. 10016. They will thus be available to qualified investigators to use in further analyses.

Chapter I

INTRODUCTION TO THE STUDY

1. Purposes

This study had two major foci:

(1) the study of foreign language teaching in the United States in so far as it affects students who "major" or "concentrate" in foreign languages at the collegiate level; and

(2) the study of the collegiate preparation of foreign language teachers.

In the conduct of the study, it was first necessary to make an assessment of the levels of foreign language competence attained by foreign language majors near the time of their graduation from college, and then to examine the student characteristics and instructional program variables associated with the various levels of measured competence. In this way it was hoped that useful conclusions could be drawn concerning the whole state of foreign language teaching in America and that certain suggestions would emerge as to ways in which instructional programs might be improved. In the case of students preparing to become teachers, attention was directed, in addition, to the degree to which these individuals had been able to attain not only the basic linguistic skills in their respective languages but also certain skills and knowledges pertaining specifically to their preparation as prospective foreign language teachers. This part of the study was therefore designed to produce useful information about the conduct of foreign language teacher preparation programs.

Ever since the establishment of its Foreign Language Program in 1953, the Modern Language Association of America has evinced much concern for the adequate preparation and certification of foreign language teachers (Axelrod, 1966; Paquette, 1966). With the support of funds from the U. S. Office of Education and with the assistance of Educational Testing Service, the MLA mounted a major project in the late 1950's to develop a series of proficiency tests in foreign language skills and teacher preparation knowledges in five languages: French, German, Italian, Russian, and Spanish (the five languages designated as "commonly taught" modern languages by the U. S. Office of Education). By 1962, this project eventuated in the MLA Proficiency Tests for Advanced Students, constructed in two forms in each language (Starr, 1961a, 1961b, 1962; Educational Testing Service, 1964a). Because these tests were carefully constructed by foreign language experts with the help of measurement specialists and laid at least as much stress on listening and speaking skills as they did on skills in reading and writing, they have come to be widely accepted in the foreign language teaching profession as valid and useful measures. Many states have adopted them for use in the certification of public school foreign language teachers, and they have been employed also by many college and university foreign language departments for testing the attainments of students at various stages of advanced training. They have received extensive use at NDEA Institutes for Foreign Language Teachers to assess the effects of institute training, and preliminary norms have been derived from the results. Recently, the administration of the tests has been undertaken by Educational Testing Service as one of its regular national testing programs.

The immediate impetus for the present study was the need for carefully collected normative information felt by states and other jurisdictions using the MLA tests for teacher certification. But as early as 1963, a conference, convened to consider problems of the preparation of college foreign language teachers, had recommended that comprehensive studies be made of the attainments of both undergraduate and graduate students as measured by the MLA tests. To quote from the conference report:

> "To enable these tests [the MLA Proficiency Tests for Teachers and Advanced Students] to produce the results of which they appear capable, the Conference recommends that a project be undertaken, with government or foundation support, to arrange for and finance the experimental administration of the MLA Tests in the four language skills, on the largest possible scale, to college majors in the relevant languages at a time near graduation and to graduate students toward the end of one or two years' residence. Definitely superior to the Graduate Record Examination in the measurement of language skills, these tests, if administered at the end of the Undergraduate Program and again during the Graduate Program, would enable language departments to measure the achievement of their students according to a recognized scale. The results of this testing would determine each student's need for additional language training in graduate school. If the tests were to be used as partial demonstration of a student's readiness for teaching, the Conference felt strongly that they should be taken again when the student proposes to begin his teaching career: the proficiency must be measured when it is to be used and not at some time in the past" (MacAllister, 1964, p.39).

Limitations of funds and other resources made it infeasible for this study to use the MLA tests in an investigation of foreign language attainments at the graduate school level, but the study has, in effect, attempted to carry out the Conference recommendations with respect to testing at the undergraduate level. In fact, in employing not only the four skill tests but also the three professional preparation tests for students intending to teach, it went beyond the Conference recommendations.

An important aspect of the present study was its attempt to calibrate the scores on the MLA Proficiency Tests in terms of quasi-absolute, inherently meaningful standards. There were, to be sure, certain tentative calibrations available from a study by Myers and Melton (1964), but ordinarily, scores on those tests could be interpreted only in relative terms, i.e., with reference to the performance of known groups such as samples of foreign language teachers attending NDEA Institutes. But percentile ranks and standard scores fail to indicate in any meaningful way the absolute amount of competence the tested individuals possess. Except to the extent that one can guess at the range of competence possessed by a reference group, a percentile rank cannot tell, for example, how successful the individual would be in communicating with a native speaker of the language or in comprehending the substance of printed materials in the language. Therefore, a substudy was designed to ascertain correspondences between MLA Proficiency Test scores and the "absolute proficiency ratings" rendered by expert teams from the Foreign Service Institute of the U. S. Department of State (Rice, 1959). These latter ratings are couched in terms of meaningful levels of proficiency and are accepted in U. S. Government circles as directly interpretable assessments of foreign language competence.

In the course of conducting a large normative study of the foreign language attainments of college foreign language majors, it was realized that much valuable subsidiary information could be collected and analyzed with regard to such questions as the following:

(a) Who are the foreign language majors? That is, what are their backgrounds and characteristics?

(b) What are their motivations for foreign language study?

(c) What kinds of exposure to foreign language study have they had? What influences have they had outside of their regular courses of study in foreign languages? To what degree are these influences associated with degree of success in foreign language study?

(d) What variations can be observed in foreign language major elections and attainments in different types of institutions or in different regions of the country?

(e) Can different types of collegiate foreign language programs be identified, and if so, are these types associated with different levels of foreign language attainments in the students?

As the principal investigator has observed (Carroll, 1963b, 1966), there is still a dearth of reliable research in foreign language teaching. Although the present study could not, by its nature, investigate foreign language learning processes and instructional procedures at close range, it was hoped that some of the information it was to collect would eventually serve as useful background for such research.

2. Overview of the Study

This study was conducted in two phases. The first phase started officially on March 15, 1964 and consisted of a pilot study in which the MLA Proficiency Tests and various other instruments were administered in the spring of 1964 to 667 individuals at 44 institutions of higher learning in the state of New York.* The purpose of this "New York pilot study" was primarily to try out procedures of gathering data and only secondarily to obtain substantive results for the sample of individuals tested. Although the study was intended to obtain data from 100% of the senior foreign language majors at 100% of the institutions of higher learning in the state of New York, the institutional response rate was only 70.1% and the overall student response rate within participating institutions was only 76.2%. As a result of experience gained in the pilot study, certain modifications were made in the procedures for gathering data with a view to obtaining higher response rates in the second phase of the study. In addition, tentative norms were established for the MLA Proficiency Tests on the basis of the sample obtained in the pilot study. A final report on this phase of the study was rendered on January 31, 1965 (Carroll, 1965); this report did not contain, however, detailed analyses of data obtained. Appendix A of the present report presents a summary of the principal results of pilot study data analyses that have been made subsequently. These data analyses are not included in the main body of this report because they are not considered as representative and significant as the analyses that have been made of the data obtained in the second phase of the study.

Planning for the second phase of the study began in the fall of 1964, although this phase did not start officially until December 1 of that year. Instruments and procedures were revised in the light of experience in the New York pilot study. A sampling design was devised on the assumption that it would be possible to obtain a 100% student response rate within a stratified pptimal sample of 192 institutions of higher learning in the whole United States such that a probability sample of 50% of all foreign language majors in the country would be tested. Contacts with institutions were made beginning on December 10, 1964; because of a disappointing institutional response rate in this first sampling, a further sample of 102 institutions was chosen and solicited. In the end, however, only 2,875 students were tested in the spring of 1965 (rather than the approximately 5,000 desired), and there were only 203 institutions that participated. The present report is mainly concerned with describing and analyzing the findings of this second, nationwide sampling study.

The substudy of MLA Proficiency Test equivalences with FSI absolute proficiency tests was done in the spring and summer of 1965, with the use of selected foreign language teachers enrolled in NDEA institutes.

The work of making the major analyses of the data collected in both the New York pilot study and the nationwide study was accomplished in the academic year 1965-66. In June, 1966, the principal results obtained up to that time were presented at a two-day invitational conference of consultants. These consultants, among whom were found both authorities on foreign language teaching and experts on educational research, made numerous suggestions as to further analyses that should be undertaken; these further analyses were made during the academic year 1966-67, which also saw the completion of the present report. One of the major concerns of analyses conducted during the academic year 1966-67 was to estimate the degree of sampling bias inherent in the results of the nationwide study; it was only during that year that it became possible to check enrollment figures against tabulations made by the statistical services of the U. S. Office of Education.

---

*The pilot study was also to include all foreign language majors graduating from Harvard University and Radcliffe College in June, 1964. However, only 4 out of approximately 50 students volunteered to take the tests.

Chapter II

MEASUREMENTS AND PROCEDURES

1. Instruments

MLA Foreign Language Proficiency Tests for Teachers and Advanced Students (Form A)

From the start of this project, there was never any doubt that the instrument of choice for the measurement of foreign language competencies at an advanced level was the series known as the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students, constructed in five languages and in two alternate forms in an extensive three-year project of the Modern Language Association, under the general direction of Wilmarth H. Starr and with the support of funds provided by the U. S. Office of Education under NDEA Title VI.

Originally conceived as "Qualifications Tests for Secondary School Teachers of Foreign Languages," the MLA tests gradually evolved into a wide-range, general battery of tests that could be used for measuring both basic foreign language skills and certain professional teaching competencies. As Starr (1962) recounts their history, a significant impetus came from a statement of "Qualifications for Secondary School Teachers of Modern Foreign Languages" formulated in 1955 by the Steering Committee of the Foreign Language Program of the MLA. These qualifications established three general levels of proficiency (Minimal, Good, and Superior) for seven areas of language teaching competencies: (1) aural understanding, (2) speaking, (3) reading, (4) writing, (5) language analysis, (6) culture, and (7) professional preparation. It became recognized, however that "the statement of desiderata, no matter how strongly representative of a consensus of the profession, would not be as effective as the situation demanded until nationally standardized tests could be developed that would implement the description of competencies." In the spring of 1959 the MLA made a contract with the U. S. Office of Education to develop such a series of tests in the five languages designated as "commonly taught", namely, French, German, Italian, Russian, and Spanish, and Starr, then Head of the Department of Foreign Languages and Classics at the University of Maine and later Head of the All-University Department of Romance and Slavic Languages and Literatures at New York University, was named Project Director. The actual work of test construction was carried out by 31 committees (one in each language in each of the first six competence areas, plus one all-language committee in the professional preparation area) composed of foreign language teachers from both secondary school and college levels. These committees were assisted by test construction experts from Educational Testing Service, which also conducted most of the work of administering the tests in field studies, and performing statistical analyses of the results. The tests went through preliminary editions before being refined, through item analyses and test critiques, into the forms that are now available from Educational Testing Service. The final standardization runs were conducted in 1961 with testing populations from 68 Summer and 7 Academic-Year NDEA Institutes for foreign language teachers. An authoritative source of information about the tests is a booklet published by ETS (1964a), A description of the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students, containing the (slightly revised) qualifications statements for Superior, Good, and Minimal levels in each area of competence, sample questions, and statistical information on reliability, means and standard deviations, and intercorrelations for the NDEA Institute norm population. Percentile norms are published by ETS in a separate leaflet (ETS, 1964b), and more detailed information is available on item statistics, etc., in various publications of restricted circulation. Myers and Melton (1964) conducted a study attempting to set up equivalences between MLA proficiency test scores and MLA-defined competence levels (Superior, Good, Minimal, and Unsatisfactory).

In an address given at the 1960 convention of the MLA, Starr (1961a) made three major points about the tests. First, he said that they "are pitched rather higher than the well known foreign language tests of the College Board, Cooperative Test, or Advanced Placement Test batteries." Second, "they are proficiency tests and as such reflect the general and growing dissatisfaction with the pattern of American education which looks upon teacher qualification quantitatively under the assumption that hours of contact with subject matter in the classroom and credits gained thereby insure the necessary qualification to teach." And third, "they are pioneering new frontiers." As far as he knew, "it is the first time the members of an academic subject-matter area have tried to identify and standardize their own competencies on a national scale."

The tests have been remarkably well received not only by the foreign language teaching profession itself but also by many state and local jurisdictions having to do with the certification or selection of foreign language teachers. At this writing, at least 16 states use them in this way, and the MLA has established an office, directed by F. André Paquette, to give advice and assistance on the use of the tests. The MLA also has a continuing contract with the U. S. Office of Education to study the use of the tests and to keep them current through the creation of further alternate forms with the cooperation of Educational Testing Service. The latter agency now conducts a regular national MLA proficiency testing program 5 times a year. In the year 1965-66, 10,600 persons were tested in various programs, national and others. In addition, the tests continue to be used in NDEA Institutes and in institutional testing programs.

The tests are not, of course, perfect, but the evaluations of them published in Buros's Sixth Mental Measurements Yearbook (1965) were in general highly favorable. The MLA has reprinted these evaluations, together with a large number of detailed evaluations it itself collected from foreign language teachers, in a document (Tollinger and Paquette, 1966) for restricted circulation.

The series consists of a battery of seven tests in each of five languages -- French, German, Italian, Russian, and Spanish. The first four tests in each battery cover the skill areas--listening, speaking, reading, and writing. The last three tests cover areas of competence important in language teaching--applied linguistics, civilization and culture, and professional preparation. Although following a common pattern and test specifications, the questions in the first six tests are different for each language; they are in no sense translations of each other. The seventh test, Professional Preparation, is entirely in English and is the same for all languages.

The seven tests in each battery are described briefly below, with approximate administration times, number of items, and maximum converted scores for Form A (the form used in the present study). Descriptions have been freely adapted from the booklet issued by ETS (1964a).

1. Listening Comprehension Test (20 minutes, 36 items, max. converted score = 56)
The student listens to a tape recording consisting of single utterances, a connected dialogue, a radio broadcast, and a three-part conversation. For all but the last part, the examinee selects, from among four choices printed in his test book (always in a foreign language), the responses most likely to be made in the situations presented or the best answers to questions put to him. For the last part, the examinee marks on his answer sheet whether each of a number of statements about the conversation he has just heard is true or false. Questions are designed to test phonetic (or phonemic) discrimination, command of idiomatic expressions, vocabulary, and structure typical for conversation use of the language. Comprehension of main facts, ideas, and important details is evaluated.

While the main object of this test is to measure listening comprehension, it depends to some extent upon reading comprehension because the alternatives to the multiple choice questions are printed in the foreign language.

2. Speaking test (15 minutes, 105 items, max. converted score = 125).
This test has three parts. The student records all his answers on tape for later evaluation by expert scorersapproved by MLA and ETS.

In Part A the student repeats 15 sentences spoken on the tape by a master voice. In Part B, he is presented with short printed texts which he must read aloud into the tape recorder. In both these parts, he is scored for the "rightness" or "wrongness" of his reproduction of certain specific sounds or intonation patterns. In addition, in Part B he is scored for the overall quality of his reading on a five-point scale.

In Part C, the examinee is presented with a single picture, a pictorial sequence, and a situation shown pictorially (with directions printed in English). In each case, he is asked to describe the picture or to react orally to the situation presented. On five-point scales, he is rated on Vocabulary, Pronunciation, Structure, and Fluency.

3. Reading test (40 minutes, 50 items, max. converted score = 70)
In Part A, the student is to select the appropriate lexical, idiomatic, or structural completion for each of 15 incomplete statements. Part B consists of several short passages followed by questions (30 in all) on the meaning of words and phrases as well as on the content of the selections. Part C contains 5 questions on short poetic excerpts. All materials--reading passages and questions--are in the foreign language.

4. Writing test (45 minutes, 60 items, max. converted score = 80).
All items in this test are of the free response type, each to be scored right or wrong by expert scorers. Part A consists of short texts in which there are, in all, 30 blanks representing omitted words that are to be filled in from the context. The student supplies, on a separate answer sheet, a single word that is appropriate both in meaning and form. (In the case of the Russian test, the base form of each word is supplied, and the student is to supply the appropriate inflected form.) Part B is an "interlinear exercise" consisting of texts "poorly written" (as if, say, by a student) containing, in all, 30 grammatical, lexical, or idiomatic errors that are to be found and corrected by the examinee. The examinee is scored right for each error properly corrected, and wrong for each error improperly corrected or unnoticed. Credit is not given for performance on parts of the test containing no planned errors. The Russian test differs slightly in that the errors in the first text are identified by underlining.)

5. Applied linguistics test (40 minutes, 55 items, max. converted score = 75).
The items cover matters of pronunciation and phonetics, orthography, morphology and syntax, and general, historical, and comparative linguistics. Questions in this test are couched in English except for cited examples in the foreign language. The general emphasis is on those differences between English and the foreign language that are of importance in teaching.

6. Civilization and culture test (30 minutes, 60 items, max. converted score = 80).
Using multiple-choice questions in English, this test endeavors to measure the knowledge of civilization and culture (as distinct from knowledge of the language) needed by a teacher to introduce his students intelligently to that area of the world in which the language is spoken. The term "culture" is taken in the broad sense, embracing such content areas as geography, history, the arts, literature, and social institutions. Within this context, questions are designed to test recall of basic facts and terms; understanding of important ideas, themes, and institutions; understanding of the complexity and variety of the culture; appreciation of historical and cultural meaning of terms; appreciation of cross-cultural differences; and understanding of relationships not only among basic facts and terms, but among ideas, themes, and institutions as well.

7. Professional preparation test (45 minutes, 65 items, max. converted score = 85).
This test is the same for all languages. It consists of multiple-choice questions designed to test not only knowledge of teaching methods that are common to different languages (with emphasis on the audio-lingual approach), but also knowledge of important

developments in the field of language testing. Questions deal with such areas as objectives, teaching techniques, professional development, and evaluation.

Form A of the MLA series was selected for the present study because, at the time the testing was done, Form B was reserved for teacher certification testing and Form C (now available for all languages except Italian) was still under development.

The Modern Language Aptitude Test, Short Form

An important student characteristic that needed to be taken account of in this survey is the student's basic aptitude for learning a foreign language. Research reviewed by Carroll (1963) has suggested that foreign language aptitude is a relatively invariant characteristic of the individual (differing among individuals, however) that is associated with the ease and rapidity with which the individual can learn a foreign language. A measure of foreign language aptitude was desired for this study in order not only to have a variable that would permit the control of various experimental comparisons but also to study further the relations between language aptitude and the attainment of competence in foreign languages under various conditions.

The Modern Language Aptitude Test (Carroll and Sapon, 1958) was selected for this purpose because it was considered to be the best test available and because, as described in the accompanying test manual, it had been extensively validated in the prediction of success in both traditional and audiolingually-oriented foreign language instruction (see also Carroll, 1962; Carroll *et al*., 1966).

The complete test, containing five parts, requires an administration time of about 60 minutes and necessitates the use of a tape recording. Parts 3, 4, and 5, however, constitute a "short form" which altogether requires only about 30 minutes and does not entail the use of a tape recording, consisting entirely of pencil-and-paper tests. Its reliability and validity are satisfactory although probably not as high as those of the complete test. To minimize testing time the "short form" was selected for use in the present study. A special answer sheet was prepared so that the test could be scored by the SCRIBE scoring machine at Educational Testing Service, along with the MLA proficiency test answer sheets.

Part 3 of the MLAT, entitled "Spelling Clues," requires the subject to recognize words "spelled approximately as they are pronounced," such as *luv* (love) and *ernst* (earnest) and choose a synonym for each word from five choices offered (*affection* and *sincere* are the correct responses for these examples). The test contains 50 items, for which a time-limit of 5 minutes is allowed, and is thus highly speeded. According to the test manual, "scores on this part depend to some extent on the student's English vocabulary knowledge"; however, it "also measures the same kind of sound-symbol association ability as measured by Part II, Phonetic Script, but to a lesser extent." [Part II is a test that must be administered by means of a tape recorder; in this test, the subject learns to associate particular English sounds with symbols in a special phonemic transcription.]

Part 4 of the MLAT, entitled "Words in Sentences," requires the subject to select elements of sentences that correspond in grammatical function to specified elements in other, "key" sentences. It can be characterized as a test of grammatical analogies. According to the test manual, "this part is thought to measure sensitivity to grammatical structure, and may be expected to have particular relevance to the student's ability to handle the grammatical aspects of a foreign language." Fifteen minutes are allowed for the 45 items of the test; for most students, the test is not highly speeded.

Part 5, entitled "Paired Associates," requires the subject to memorize the English meanings of 24 nonsense words labeled as "Kurdish" (they are not really Kurdish); 4 minutes are allowed for this memorization, after which there is a 4-minute, non-speeded multiple-choice test on the memorization. Maximum score is 24.

Questionnaire for Foreign Language Majors

It was deemed desirable to collect from each student in the testing program a wide variety of information concerning his background, interests, and attitudes in the study of foreign languages (particularly, the language in which he was "majoring"). Such information was to be studied in relation to his test scores.

The Questionnaire for Foreign Language Majors used in the national testing program was a revision and adaptation of a Questionnaire for Students Participating in the Pilot Program that had been developed for the New York pilot study. Refinement consisted primarily of editing phraseology and converting certain questions from a free-response to a multiple-choice form. The final form of the questionnaire is to be found in Appendix B.

Essentially, it has three parts. The first page asks for the student's name, the permanent address to which the test results could be mailed, sex and age, expected date of graduation and institution, language in which he would be tested and data on any other languages included in his official "major subject," years of study had in other languages, and whether he had ever taken the MLA Proficiency Tests before. The second part is contained on pages 2 and 3 and asks for a detailed (but student-coded) history of the courses he has taken in his major language, including information on courses at "grade school level," high school, and college, the duration of the courses, class hours per week, size of class, teacher's classroom language, student's classroom language, use of language laboratory, quality of teacher's pronunciation, use of native informants, whether the final examinations involved listening and speaking components, and the student's final grade. The third part, on page 4, explored the reasons the student majored in a foreign language, his possible intentions to teach and if so at what levels, his ratings of the relative importance of listening, speaking, reading, and writing skills to him, plans to qualify for a teaching certificate in foreign languages, informal courses taken in the major language, courses taken in Latin or Greek, travel abroad in the country where the major language is spoken, parents' use of the major language at home, other opportunities to use the language, and the amount of reading experience in the major language.

Information obtained with this questionnaire was mainly of an objective nature; the ways in which the answers were coded and summarized will be set forth in later chapters of this report.

Questionnaire for Chairmen of Foreign Language Departments

In order to gather data on collegiate instructional and teacher training programs that might be found relevant to the interpretation of the results of the proficiency testing, a Questionnaire for Chairmen of Foreign Language Departments was used (Appendix C). Except for a minor change (correcting German Review to Germanic Review) this was identical to the questionnaire of the same name used in the New York pilot study, since the pilot study had not disclosed any other suggestions for revising it. This form asked for information on:

Languages in which courses are given in the department
Time at which students normally elect majors
Guidance given to students with regard to election of foreign language majors
Requirements for acceptance of students as foreign language majors
Degree requirements for foreign language majors
Degree programs offered beyond the A.B.
Number of majors enrolled in the department, in each language
Relevant professional periodical holdings in the institutional or departmental library
Types of foreign language courses offered
Special activities offered in the foreign language program (e.g. study abroad, language houses, language clubs, etc.)
Policies in teaching the basic skill courses

Use of language laboratory
Amount of use of English in foreign language courses

A considerable amount of the information obtained in this questionnaire was of a free-response nature and had to be hand-coded and summarized in a series of statistical variables.

2. The establishment of tentative MLA proficiency test equivalences with FSI (Foreign Service Institute) absolute proficiency ratings

Although one of the important objectives of the study was to establish norms on the MLA Proficiency Tests for foreign language majors in U. S. colleges and universities, it was realized that norms, in themselves, fail to give meaningful information as to the absolute levels of proficiency represented by given scores. For example, knowing that a student attains a score corresponding to the 50th percentile on a test tells one only that he is at the median of the group on which the test was standardized. In the case of a foreign language test, it does not indicate how much competence the individual has attained--it does not say, for instance, how competently the individual could read a foreign language newspaper or how fluently he could give a lecture in the foreign language.

The study by Myers and Melton (1964) attempted to produce meaningful ranges of score levels on the MLA Proficiency Tests in terms of the adjectival standards (Superior, Good, Minimal, Unsatisfactory) that had been set by the Modern Language Association in its 1965 statement of "Qualifications for Secondary School Teachers of Foreign Languages," or more strictly, in terms of the verbal descriptions of those standards (e.g. for Superior Listening Comprehension, "Ability to follow closely and with ease all types of standard speech, such as rapid or group conversation and mechanically transmitted speech"). The technique was to establish equivalences on the basis of qualitative ratings given to approximately 3,000 teachers who attended NDEA Foreign Language Institutes during the summer of 1963. These ratings were assigned for each of the seven areas of competence (Listening, Speaking, etc.) presumably measured by the tests. According to Myers and Melton, "in almost all cases the ratings were given by faculty groups rather than by individual faculty members," and they imply that the ratings were made without knowledge of the students' test scores, at least, not those on Form B of the MLA Proficiency Tests administered at the conclusion of the instructional program at each institute. When the test scores became available, they were compared with the ratings in the following way:

> "The procedure used for determining the score ranges was that of dividing the score distributions into four groups in such a way that the per cent in each group would be as close as possible to the per cent in the comparable rating category. For example, 214 individuals were rated Superior for French Listening Comprehension, approximately 17 per cent of the 1,279 for whom test scores and ratings were available. In this same group, 211 persons had scores of 52 or higher (while 260 had scores of 51 or higher). Therefore 42 to 58 was designated as the score range corresponding to the category of Superior...All other score ranges were identified in the same way except in the case of the Russian tests. Since there were not enough participants at the Russian institutes to allow for a distinction between the two lowest categories, only three score ranges are reported."

They further comment:

> "The score ranges...should not be viewed as absolute or definitive. In the first place, the correlations between the test scores and the corresponding ratings were not

perfect (and should not be expected to be perfect). [Actually, for skills tests the correlations between corresponding test scores and ratings were as follows: French, .63 to .72; German, .67 to .74; Russian, .49 to .68; Spanish, .62 to .72. For the three teacher preparation tests the correlations were generally lower: French, .36 to .55; German, .29 to .60; Russian, .19 to .36; Spanish, .30 to .43--Author's note.] Hence, it cannot be said that a person whose test score is in the Superior range would necessarily be rated Superior in that skill by an Institute staff. Furthermore, neither the score nor the rating was designed to be a predictor of the other; both the test score and the rating were designed to be estimates of teaching competence." [This statement, it should be noted, is debatable; as far as the skill areas are concerned, one would think that the test scores and the ratings were designed to measure aspects of foreign language competence.—Author's note.]

Above all, however, one could question the validity of the ratings, since they were based on general impressions of performance rather than upon standardized procedures of assessment. It therefore seemed that a new attempt should be made to establish meaningful equivalences to absolute standards, at least for the four skill areas.

Since the late 1950's, the Foreign Service Institute of the U. S. Department of State has used a standardized interview procedure to render what are called "absolute language proficiency ratings" on two scales: S (Speaking Proficiency), and R (Reading Proficiency). The procedure, as described by Rice (1959), is applicable to any language, and is conducted by trained teams composed of a linguist and a native speaker of the language. Interviews are conducted on an individual basis and take about 30 to 45 minutes. The ratings are rendered on essentially a six-point scale, from 0 to 5, except that all the ratings except 5 "may be modified by a plus (+), indicating that proficiency substantially exceeds the minimum requirements for the level involved but falls short of those for the next higher level" (Foreign Service Institute, 1963). (Thus, the scale may be regarded as an 11-point scale for statistical purposes; i.e., 0, 0+, 1, 1+, etc.) Five points on the scale (1, 2, 3, 4, 5) are accurately described both in a short definition (see bottom of Table 2.2) and an amplified statement available in a Foreign Service Institute Circular (Foreign Service Institute, 1963). Although the two scales are named Speaking and Reading, the descriptions also include some reference to listening skills (for the Speaking Proficiency scale) and to writing skills (for the Reading Proficiency scale). The meanings of the scale points are widely accepted and apparently well understood in U. S. government circles.

It was therefore decided to attempt to establish equivalences between scores on the skills tests of the MLA Proficiency Tests and the FSI absolute proficiency ratings, at least for French, German, Russian, and Spanish.

The original plans for this equivalency study called for the obtaining of 50 cases in each language, using students already tested in the national study reported here. Accordingly, arrangements were made to send FSI interview teams to selected institutions in Eastern U. S. already participating in the study. But the project staff failed to reckon on the fact that the regular MLA Proficiency Testing program had already been a sufficient imposition on the time and patience of both the staffs and students at those participating institutions. There was only one institution (Queens College) where it became possible to arrange for FSI testing, and even here, only 6 cases, all in Spanish, were obtained (in May, 1965).

To obtain the necessary cases, the project staff considered the possibility of obtaining cooperation from selected NDEA Summer Institutes for Foreign Language Teachers where the MLA Proficiency Tests would be given both at the outset of the program and its conclusion. If FSI teams could be sent to such institutes and give their interviews to selected students, it would be easy to compare the FSI ratings with both pretest and posttest MLA test scores and from such data to establish equivalences. As a matter of fact, through fortunate circumstances, an FSI team that had attempted but failed to arrange for testing of senior majors at Indiana University (one of the institutions participating in the study) was nevertheless able to arrange for testing, in May 1965, of a group of 24 Russian teachers in an Academic Year NDEA Institute at that institution.

Accordingly, on June 24, 1965, a letter was addressed to the directors of ten NDEA summer institutes asking whether they would permit small numbers of their students to be tested by FSI teams. Each institute was given a quota of 20 students, and the institute directors were told that the students should represent "the broadest possible spectrum of ability in the language in which they will be tested." They were asked to select for each language "four of your ablest students, four of your 'good' students, four of your average students, four of your below average students, and four of your least proficient students." As it turned out, it became possible to do FSI testing at five institutes: 40 cases were obtained in French at Rutgers and the University of Pittsburgh; 30 cases in Spanish at Rutgers and Gannon; and 39 cases in German at Princeton and Hofstra. Together with the 6 student cases in Spanish obtained at Queens and the 24 Russian cases at the Indiana University Academic Year Institute, this made a total of 128 cases. Table 2.1 shows the number of cases in each language and the range of FSI ratings represented.

Table 2.1

Cases Tested by FSI and Ranges of FSI Ratings

Language	N	Ranges of FSI Ratings S ratings	 R ratings
French	40	1 to 4	2 to 4+
German	39	1+ to 5	1 to 5
Russian	24	1 to 4	1 to 3+
Spanish	36	1+ to 4	1+ to 4
Total	128		

Arrangements were made with Educational Testing Service to retrieve and transmit to the project staff the MLA Proficiency Test scores for all persons involved in this study. The MLA Proficiency Test scores used in subsequent analyses were the "post-institute" scores, i.e., those resulting from tests administered toward the end of the institute period (whether it was an Academic Year Institute or a Summer Institute). For the six Spanish cases derived from the college senior testing, the MLA Proficiency Test scores were those obtained in that testing.* There were a few cases where for some reason the

---

*The six Queens College senior cases in Spanish were not used in the analyses, however, because the MLA Speaking test scores and FSI ratings were clearly off the line of equivalence for the 30 NDEA cases. The Queens College seniors' MLA Speaking Scores were all much lower than one would expect them to be from the FSI ratings they received and in fact showed no relationship with the FSI ratings. Apparently the scoring standards used in judging the tapes were different for the NDEA Institute and national study cases. This finding is disturbing and casts doubt on all equivalences for MLA Speaking scores and indeed upon the interpretation of those scores. The Queens College cases could, nevertheless, have been pooled with the NDEA Institute cases in the derivation of equivalents for Listening, Reading, and Writing scores; the correlations between corresponding tests and FSI ratings when all 36 cases were used were actually higher than for the NDEA Institute cases alone. However, to preserve uniformity in the data it was decided not to include the Queens College cases in the equivalency analyses for any scores.

MLA Proficiency Test scores were not available.

Table 2.2 shows a statistical summary of the basic data obtained in this FSI-MLA equivalency study.

An immediate question is whether the FSI-MLA equivalency groups can be regarded as representative of the groups tested in the national college senior study whose results are to be presented in Chapters III to IX of this report. Statistical comparisons with results from the national study (Table 4.1 , Chapter IV) show the following: The French group is significantly superior (at the 1% level) to the national group in Listening and Speaking, but not in Reading or Writing; the German group is significantly superior (at the 1% level) to the national group only in Speaking; the Russian group is representative of the national group in Listening, significantly superior (1% level) in Speaking, and significantly inferior in Reading and Writing; the Spanish group is representative of the national group in all skills except in Reading, where it is significantly inferior (1% level). There were only two comparisons (German, Speaking; Spanish, Reading) where the variances of the FSI group were significantly (1% level) unrepresentative of the variances in the national study. On the whole, the FSI equivalency groups are not markedly unrepresentative of the national study groups.

The mean levels of FSI ratings were in the neighborhood of 2+ and 3, that is, in the neighborhood of "minimal professional proficiency", except in the case of Russian, where the mean ratings were in the neighborhood of 2, "limited working proficiency." It would have been desirable to have obtained a Russian group with higher average proficiency.

Correlations between FSI ratings and corresponding MLA scores were quite high, even in the Russian group. They compare favorably with the correlations between ratings and corresponding skills scores obtained in the Myers and Melton study (see p. 11). (In the present study, FSI Speaking ratings are regarded as corresponding to both Listening and Speaking scores on the MLA tests, and the FSI Reading ratings are regarded as corresponding to both Reading and Writing scores on the MLA tests.) The correlations between the two FSI ratings, S and R, are quite high, ranging from .69 for French to .90 for Russian. Unfortunately, save possibly in the case of French, there is little evidence in the FSI-MLA correlations to suggest that FSI Speaking ratings are more highly correlated with MLA Listening and Speaking scores than with MLA Reading and Writing scores, nor that FSI Reading ratings are more correlated with MLA Reading and Writing scores than they are with Listening and Speaking scores. Nevertheless, on an *a priori* basis, it was decided to establish FSI-MLA equivalences using the correspondences defined above. Scatterplots of corresponding FSI ratings and MLA scores revealed no evidence of curvilinearity of relationship.

Equivalences were established by the "equal standard scores" method. That is, using the means and standard deviations of corresponding distributions, linear equations were set up such that corresponding values specified by the equation would have the same standard score. (A standard score is obtained by finding the algebraic difference between a score and the mean and dividing it by the standard deviation.) Thus, if X is the FSI rating and Y is the corresponding score on an MLA test, the equal standard scores are respectively the left and right sides of the equation

$$\frac{X - \bar{X}}{\sigma_X} = \frac{Y - \bar{Y}}{\sigma_Y}$$

This equation can, of course, be solved explicitly either for X or for Y. Solving for Y, we have

$$\begin{aligned} Y &= (Y - \frac{\sigma_y}{\sigma_x}\bar{X}) + \frac{\sigma_y}{\sigma_x} X \\ &= a + bx. \end{aligned}$$

Table 2.2

Basic Data for Equivalency of FSI "Absolute Proficiency Ratings" and Scores on MLA Proficiency Tests for Teachers and Advanced Students

	FRENCH		GERMAN		RUSSIAN		SPANISH	
No. tested	39		39		19		30*	
Test	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
MLA List.	47.38	6.07	45.62	7.93	41.84	5.48	44.57	5.71
" Speak.	82.97	9.84	97.90	19.83	85.00	10.92	84.87	9.49
" Read.	50.41	7.82	51.59	11.03	33.16	8.96	43.77	5.94
" Write	49.05	8.10	55.51	14.26	56.32	9.31	52.83	10.72
FSI Speak.*	2.62	.64	3.13	1.08	1.97	.66	2.58	.75
FSI Read.*	3.15	.66	3.10	1.10	1.89	.57	2.86	.66

Correlations with FSI Ratings

	"S"	"R"	"S"	"R"	"S"	"R"	"S"	"R"
MLA List.	.67	(.61)	.73	(.72)	.84	(.75)	.73	(.80)
" Speak.	.67	(.49)	.82	(.83)	.78	(.66)	.66	(.65)
" Read.	(.58)	.71	(.82)	.82	(.78)	.69	(.63)	.74
" Write	(.65)	.63	(.86)	.84	(.62)	.71	(.70)	.77
FSI "R"	(.69)		(.95)		(.90)		(.80)	

* In computing these values, a "+" is given a value of .5. Thus, 1+ is coded 1.5, 2+ = 2.5, etc. For the meanings of the FSI ratings, see below.

Native or bilingual proficiency	S-5 Speaking proficiency equivalent to that of an educated native speaker.	R-5 Reading proficiency equivalent to that of an educated native speaker.
Full professional proficiency	S-4 Able to use the language fluently and accurately on all levels normally pertinent to professional needs.	R-4 Able to read all styles and forms of the language pertinent to professional needs.
Mimimum professional proficiency	S-3 Able to speak the language with sufficient structural accuracy and vocabulary to satisfy representation requirements and handle professional discussions within a special field.	R-3 Able to read non-technical news items or technical writing in a special field.
Limited working proficiency	S-2 Able to satisfy routine social demands and limited office requirements.	R-2 Able to read intermediate lesson material or simple colloquial texts.
Elementary proficiency	S-1 Able to satisfy routine travel needs and mimimum courtesy requirements.	R-1 Able to read elementary lesson material or common public signs.

"All the ratings except the S-5 and R-5 may be modified by a plus (+), indicating that proficiency substantially exceeds the mimimum requirements for the level involved but falls short of those for the next higher level."

—Extracted from "Absolute Language Proficiency Ratings,"
Circular, May 1963, Foreign Service Institute,
Washington, D.C.

* The 6 Queens College senior cases were not used in this analysis.

It should be noted that this equation does not specify a regression line; that is, it cannot be used to make optimal predictions of X from Y or Y from X. It merely assumes that X and Y are equally estimates of the same thing and that it is an arbitrary matter whether one measures this thing by X or by Y. Conceptually, this method of equating is similar to that employed by Melton and Myers (1964). The more X and Y are correlated, the more this procedure is justified. It is felt that in the present case, the corresponding measurements are sufficiently well correlated to justify the procedure, particularly in view of the fact that the purpose of the study was merely to establish meaningful standards for the interpretation of the MLA scores.

Table 2.3 shows the MLA equivalents (in "converted scores") for given FSI rating levels for the four tests in French, German, Russian, and Spanish; it also shows the linear equations Y = f(X) and X = f(Y).

It will be noted that in a number of cases, the corresponding MLA scores exceed the maximum possible scores. This finding would suggest that the MLA tests in those cases do not have a high enough "ceiling," that is, that they do not have the capacity to discriminate among the upper levels of FSI ratings or indeed among the upper levels of language competence (near-native and native language ability). This matter will be examined again when the distributions of MLA test scores from the national study are inspected.

Table 2.3 also shows the ranges of Superior, Good, Minimal, and Unsatisfactory rating equivalences established by Myers and Melton (1964). It will be seen that in general in Listening, Minimal corresponds to S1 and S1+, Good to S2 to S3, and Superior to S3+ and above; in Speaking, Minimal extends from S1 to S2, Good from S2+ to S3+, and Superior to S4 and above; and in Reading and also in Writing, Unsatisfactory corresponds to S1, Minimal to S1+ to S2, Good from S2+ to S3+, and Superior to S4 and above. There are some variations across languages, however: for example, in Russian, either the FSI ratings are unduly severe or the Myers-Melton ratings are unduly generous. It is largely a matter of opinion whether the FSI equivalences established here are more reliable than the Myers-Melton ratings. In favor of the FSI equivalences is the fact that they were based on a highly standardized procedure involving normal two-way communication between examiner and examinee; in favor of the Myers-Melton equivalences is the fact that they were established on the basis of many more cases than were obtained for the FSI equivalences. Strictly speaking, the Myers-Melton ratings apply only to scores on Form B of the MLA tests, but since converted scores were established to provide equivalences between Forms A and B, the Myers-Melton ratings should also apply to the Form A scores.

3. The Design of the Sample

There were several reasons for desiring a large and representative sample of college senior foreign language majors graduating at the end of the 1964-65 academic year. A large sample was desired not only because such a sample would yield results that the layman in statistics would be more likely to regard as convincing, but also because if it were desired to break the sample down into various subgroups for special analyses, such subgroups would be larger than might otherwise be the case and hence would be more likely to produce statistically significant findings. The sample had to be representative in order to serve as an appropriate basis for the development of norms that could be relied upon in the interpretation of scores in future administrations of the tests.

In the original plan for the project, written early in 1964, the assumption had been made, based on projections from enrollment data then available, that the total number of students graduating at the end of the academic year 1964-65 would exceed 10,000 and that a representative sample of 7000 distributed proportionately among the five languages would be large enough to be persuasive. After the New York pilot study

Table 2.3

MLA Proficiency Test Equivalents (Y) and Parameters of Linear Equations Relating Them to FSI Ratings (X)[1]; with Ranges of Myers and Melton Ratings[2]

FRENCH

FSI Levels[1]	MLA List.	MLA Speak.
S1	32.2 M	58.3
S1+	36.9	66.0 M
S2	41.6	73.6
S2+	46.3 G	81.2
S3	51.0	88.8 G
S3+	55.7	96.5
S4	60.4*S	104.1
S4+	65.2*	111.7 S
S5	69.8*	119.4

		List.	Speak.
Y=a+bX	a	22.53	42.69
	b	9.48	15.38
X=a'+b'Y	a'	-2.374	-2.773
	b'	.105	.065

FSI Levels	MLA Read	MLA Write
R1	24.9 U	22.6 U
R1+	30.8 M	28.7
R2	36.7	34.9 M
R2+	42.6	41.0
R3	48.6 G	47.1
R3+	54.5	53.3 G
R4	60.4	59.4
R4+	66.3 S	65.6 S
R5	72.3*	71.7

	Read	Write
a	13.09	10.39
b	11.85	12.27
a'	-1.105	-0.848
b'	.084	.082

GERMAN

FSI Levels	MLA List.	MLA Speak.
S1	30.0 M	59.0 U
S1+	33.7	68.1
S2	37.4	77.2 M
S2+	41.0 G	86.4
S3	44.7	95.5 G
S3+	48.3	104.7
S4	52.0 S	113.8
S4+	55.7	122.9 S
S5	59.3*	132.1*

		List.	Speak.
Y=a+bX	a	22.64	40.43
	b	7.34	18.36
X=a'+b'Y	a'	-3.083	-2.206
	b'	.136	.054

FSI Levels	MLA Read	MLA Write
R1	30.5 U	28.2 U
R1+	35.5 M	34.7 M
R2	40.5	41.2
R2+	45.5	47.7
R3	50.5 G	54.1 G
R3+	55.6	60.6
R4	60.6	67.1
R4+	65.6 S	73.6 S
R5	70.6*	80.1*

	Read	Write
a	20.51	15.32
b	10.03	12.96
a'	-2.043	-1.180
b'	.100	.077

Table 2.3 continued

RUSSIAN

FSI Levels		MLA List.	MLA Speak.
S1		33.7 M	68.9 M
S1+		37.9	77.1
S2		42.0 G	85.4 G
S2+		46.2	93.7
S3		50.4	102.0
S3+		54.5	110.2 S
S4		58.6*S	118.5
S4+		62.8*	126.8*
S5		67.0*	135.0*
Y = a + bX	a	25.48	52.41
	b	8.30	16.55
X = a'+b'Y	a'	-3.067	-3.164
	b'	.120	.060

FSI Levels		MLA Read	MLA Write
R1		19.2 M	41.8 M
R1+		27.0	49.9
R2		34.8 G	58.0 G
R2+		42.6	66.1
R3		50.4	74.2
R3+		58.2 S	82.3*
R4		66.0	90.4*S
R4+		73.8*	98.5*
R5		81.5*	106.6*
	a	3.45	25.45
	b	15.72	16.33
	a'	-0.219	-1.557
	b'	.064	.061

SPANISH

FSI Levels		MLA List.	MLA Speak.
S1		32.6	64.9
S1+		36.4 M	71.2 M
S2		40.1	77.5
S2+		43.9 G	83.8 G
S3		47.7	90.1
S3+		51.5	96.4
S4		55.3 S	102.6 S
S4+		59.0*	108.9
S5		62.8*	115.2
Y = a + bX	a	24.93	52.22
	b	7.61	12.65
X = a'+b'Y	a'	-3.272	-4.125
	b'	.131	.079

FSI Levels		MLA Read	MLA Write
R1		26.8 U	22.3 U
R1+		31.4 M	30.5
R2		35.9	38.7
R2+		40.5	46.8 M
R3		45.0 G	55.0 G
R3+		49.5	63.2
R4		54.1	71.4
R4+		58.6 S	79.6 S
R5		63.1	87.8*
	a	18.03	6.38
	b	9.00	16.24
	a'	-2.003	-0.394
	b'	.111	.062

*Exceeds maximum possible score

[1] For computational purposes, a "+" is given a value of .5; thus, 1+ is coded 1.5, 2+ = 2.5, etc.

[2] U = Unsatisfactory; M = Minimal; G = Good; S = Superior

testing was done, a proposal for the national study made in November, 1964 stated that in view of budgetary considerations a 50% sample, i.e. 5000 students, would be more practical. The final design of the sample was in any case not attempted until November, 1964, when figures on 1962-63 college enrollments became available from the U. S. Office of Education making projections to 1964-65 more viable. At that time it was estimated that the 1964-65 numbers of graduating foreign language majors would be approximately 12,700. This estimate was made on the assumption of a continuing and more or less constant growth rate evident in the following figures:

Year	FL Earned Degrees	Annual Growth Ratio
1960-61	5451 (actual)	- -
1961-62	6714 (actual)	1.232
1962-63	8334 (actual)	1.241
1963-64	10,292 (projected from a growth rate of 1.235)	
1964-65	12,711 (projected from a growth rate of 1.235)	

In the belief that this projection was perhaps too optimistic, a slightly more conservative figure of 12,000 was taken as the measure of the target population of the study. (This estimate, it now turns out, was remarkably accurate. Official figures from the U. S. Office of Education show 12,220 as the total number of 1964-65 graduates in the five languages.) Because previous decisions had indicated the choice of 50% as the overall sampling fraction, it was decided to design a representative sample containing 6000 students to be tested. The budgeting implications of this decision were approved by the sponsors of the project.

The sampling design was developed in consultation with Dr. Theodore Colton, a specialist in sampling theory who has been associated with Professor W. G. Cochran of the Harvard University Department of Statistics and who was on the staff of the Department of Preventive Medicine in the Harvard Medical School. Basically, the design called for stratifying the total population by "size" of institution (i.e., by the total number of students majoring in the five test languages, as shown in figures for 1962-63). An optimal set of sampling fractions for the several size-strata were developed in view of (a) the cost of testing at an institution of a given size, (b) the variance of mean test scores within a stratum, and (c) the total budget allocated to testing. Sampling within each stratum was random, except that there was a control for the type of institution (public vs. private), these two types of institutions being proportionately represented. Variances of means within size strata were estimated on the basis of data obtained in the New York pilot study (and found to be approximately equal). Relevant discussion and formulas pertaining to the sampling design are to be found in Cochran (1963, p.95f, 289) and Hansen, Hurwitz, and Madow (1953, pp. 220f.)

"Public" institutions are defined as those supported and administered under the jurisdictions of state or city authorities e.g. University of California, University of Pittsburgh; "private" institutions are those independent of such jurisdictions, dependent mainly on endowments or private funds, e.g. Princeton University, Goucher College.

Actually, the sample was designed in several stages because of a number of complications: (a) figures on institutional sizes were in terms of 1962-63 data and would have to be re-adjusted to allow for projected institutional growth; and (2) after preliminary returns were received from the first "wave" of invitations to participate, it was possible to make more accurate projections of the 1964-65 figures (since the institutions reported their current enrollments), but at the same time it developed that the institutional response rate would be less than hoped for and would have to be adjusted for by the drawing of a supplementary sample.

Five "size" categories of institutions were established somewhat arbitrarily in such a way that there would be roughly equal numbers of students in the total population in each. The basic data available for designing the sample are shown in Table 2.4, the number of public and private institutions in the five categories, and the number of students who earned degrees in each. The sampling design did not involve taking account of the numbers of students in the separate languages, it being assumed that random factors would result in an approximately proportional distribution, over the five languages, in any properly chosen sample.

Table 2.4

1962-63 Population of U. S. Institutions
Offering A. B. Degree Programs in
Test Languages, with Numbers of Earned Degrees
Source: U. S. Office of Education

Size category (No. of students in all languages)	Stratum	Public Institutions		Private Institutions		Totals	
		No. of Institutions	No. of Earned Degrees	No. of Institutions	No. of Earned Degrees	No. of Institutions	No. of Earned Degrees
40 and over	5	25	1881	11	674	36	2555
20-39	4	27	746	43	1122	70	1868
10-19	3	45	614	93	1240	138	1854
5-9	2	61	418	134	895	195	1313
1-4	1	75	179	257	575	332	754
Totals		233	3838	538	4506	771	8344

First sampling allocation

The goal of the sampling design, in every case, was to determine sampling rates for each size stratum such as to minimize the error in estimating a population mean of a test variable, under a given budgetary restriction and under given assumptions as to the cost of testing. A first determination of sampling fractions was made on the assumption that if the study had been done in 1963, a 60% sample might have been drawn containing 5000 students, and that such a study would cost \$17 per student plus \$25,000 for institutional and administrative costs, or a total of approximately \$110,000 a figure that was in the neighborhood of the amount budgeted for testing. It was thought that with optimal sampling allocation, the institution costs might be reduced. It was planned also, after this sampling allocation, to re-scale the sample to allow for projected 1964-65 figures.

First, a cost function was established as

$$\$110{,}000 = \sum^{5} (C_h + \$17\ M_h)\ n_h$$

where C_h = administrative cost for obtaining each institution in the h^{th} stratum, M_h = average number of senior FL majors per institution in the h^{th} stratum, and n_h = number of institutions in the sample of the h^{th} stratum, there being five strata. It was assumed that the second-stage sampling rate would be 100%, i.e. that all senior FL majors within a given institution would participate in the testing. Thus, the sampling design may be regarded as calling for a "clustered" sample.

The solution for values of n_h such that the variance of a mean test score is minimized subject to the total budget of \$110,000 is given by

$$n_h = \frac{W_h S_h}{\sqrt{C_h + \$17 M_h}} \Bigg/ \frac{\Sigma\left(W_h S_h \sqrt{C_{ih} + \$17 M_h}\right)}{\$110{,}000} ,$$

where

$$W_h = \frac{N_h M_h}{\Sigma(N_h M_h)} = \text{proportion of students in the } h^{th} \text{ stratum,}$$

$S_h{}^2$ = Variance between means of schools in the h^{th} stratum, and

N_h = total number of institutions in the h^{th} stratum.

If we assume that S_h^2 is constant for all h, we get

$$n_h = \frac{W_h}{\sqrt{C_h + \$17\ M_h}} \Bigg/ \frac{\Sigma\left(W_h \sqrt{C_h + \$17 M_h}\right)}{\$110{,}000} .$$

Table 2.5 is a work-sheet showing the computations starting from the data on total numbers of institutions and students given in Table 2.4. The administrative costs per school, C_h, were computed on the basis of the fact that one coordinator, to be paid \$30, would be needed even for the smallest school, and for schools with more than 20 students to test, varying numbers of proctors in addition to the coordinator would be needed. A preliminary result attained in column 10 is that the number of institutions to be selected in stratum 5 (the large-size institutions) was more than the number of institutions available in this stratum. Therefore, the sampling fraction for this stratum is taken to be 100%. This stratum would cost \$(62 x 36 + 17 x 2555) = \$45,667,

Table 2.5

Worksheet for First Sampling Allocation

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
										Final Allocation	
Stratum (h)	"Size" (No. of FL Majors, '62-'63, per Institution)	No. of Institutions (from Table 2.4) (N_h)	No. of Students (from Table 2.4) (N_hM_h)	Proportion of Students (W_h)	Aver. No. Students per Institution (M_h)	Cost per Institution (C_h)	$\sqrt{C_h + 17M_h}$	$\frac{W_h}{\sqrt{C_h + 17M_h}}$	$\frac{W_h\sqrt{C_h + 17M_h}}{A}$ (n_h)	Institution n_h	Students n_hM_h
5	40+	36	2555	.3062	70.97	62	35.61	.008599	43 (>36)	36	2555
4	20-39	70	1868	.2239	26.69	40	22.22	.010076	50	58	1548
3	10-19	138	1854	.2221	13.43	30	16.07	.013821	69	80	1074
2	5-9	195	1313	.1574	6.73	30	12.02	.013095	65	75	505
1	1-4	332	754	.0904	2.27	30	8.28	.010918	54	63	143
		771	8344	1.0000						312	5825

$$A = \frac{\Sigma W_h\sqrt{C_h + 17M_h}}{\$110{,}000} = \frac{22.088447}{\$110{,}000} = .0002008$$

Re-allocation of Strata 1-4

(13)	(14)	(15)	(16)
h	Proportion of Students (W_h)	$\frac{W_h}{\sqrt{C_h + 17M_h}}$	n_h
4	.3227	.014523	58
3	.3203	.019932	80
2	.2268	.018869	75
1	.1302	.015725	63
	1.0000		

Values transferred to column (11)

$$A = \frac{16.120200}{\$64{,}333} = .002506$$

leaving $64,333 as the total budget remaining. The remainder of the worksheet is a reallocation of the other four strata, giving the final sampling fractions shown in the last column.

It was noted, however, that while the optimal allocation procedure would yield a sample of 5825 students, it required testing at 312 institutions and did not take into account the growth of student bodies that would undoubtedly have occurred in those institutions. Basing the sample on these computations would risk getting too many students--more than our budget would allow for. Consequently, a further set of computations was performed (Table 2.6) on the assumption that all student bodies (i.e., senior FL majors) would in 1964-65 be 156% of the figures for 1962-63. This was arrived at on the assumption of an annual growth ratio of 1.25 applied for two years. Costs per school were refigured to allow for population growth. The resulting allocation required a total of 188 institutions, with sampling rates as shown in column 12 of Table 2.6. Computations showed that the estimated number of students to be tested according to this sampling plan would be 6014, at an estimated cost of $109,993, a figure close to the budgeted figure. Because this sampling plan was based on a liberal estimate of population growth, it was actually conservative with respect to the budget because it protected us from having to test an unexpected number of students at too large a number of institutions.

The next step was to apply the sampling rates to the numbers of institutions in Table 2.4 in such a way that "public" and "private" institutions would be proportionately represented in each stratum. Because the sampling rates were adjusted to actual fractions (4/5, 1/2, etc. in col. (13) of Table 2.6) to facilitate the use of random number tables in the drawing of the sample of institutions from lists, the final numbers of institutions in the various strata, given in the "number invited" column of Table 2.7, are not always exactly equal to the numbers specified in the allocation computations. It was on the basis of Table 2.7 that a mailing was made on December 10, 1964 to 192 institutions inviting them to participate in the study. In 1962-63, the total number of FL majors in these 192 institutions was 3,869. Using the previously explained growth adjustment of 156%, it was estimated that there would be approximately 6,000 students in the sample in 1964-65 if all institutions and all students responded. This number suited both the plans and the budget.

Second sampling allocation

On January 17, 1965, the principal investigator felt that it was necessary to reconsider the sampling design in view of indications that the institutional response rate would not be as high as had been hoped for. As of January 14, replies had been received from 147 (76.6%) of the 192 institutions originally invited to participate (see Table 2.7). Of the 147 institutions heard from, 124 (84.4%) responded favorably. (One of these stated that some but not all of its foreign language departments would participate.) The 124 institutions responding favorably constituted 64.6% of the 192 institutions invited. The total actual number of students reported by the favorably-responding institutions was 3678, or 61.3% of the 6000 students projected for the sample. By comparing this number with the projected number (4040) we would have expected at these institutions on the basis of the projection ratio of 1.56, a new estimate of the actual two-year growth was possible, namely 1.42.

A redesign of the total sample, using the new growth ratio of 1.42, was thus possible. It was assumed that all institutions invited would eventually be heard from, prompted if necessary by reminder letters and telephone calls. It was noted, however, that the rate of favorable reply among institutions varied according to strata, but not, apparently, according to the public-private variable. Small institutions were less likely to reply favorably than large institutions. Account was taken of this variation by appropriate adjustments in sampling rates after the reallocation computations shown in Table 2.8.

Further, all sampling rates were multiplied by 1.30 in order to allow for the non-reply rate of 77% established on the basis of the number of replies received as of

Table 2.6

Worksheet for First Sampling Allocation Adjusted for Projected Growth

(1) Stratum (h)	(2) "Size" (No. of FL Majors '62-'63) per Institution	(3) No. of Institutions (from Table 2.4) (N_h)	(4) 156% of No. of Students in Table 2.4 (N_hM_h)	(5) Proportion of Students (W_h)	(6) Aver. No. Students per Institution (M_h)	(7) Refigured Cost per Institution (C_h)	(8) $\sqrt{C_h + 17M_h}$	(9) $\frac{W_h}{\sqrt{C_h + 17M_h}}$
5	40+	36	3986	.3062	110.71	81	44.30	.006912
4	20-39	70	2914	.2239	41.63	44	27.42	.008166
3	10-19	138	2892	.2221	20.95	35	19.78	.011229
2	5-9	195	2048	.1574	10.50	30	14.44	.010900
1	0-4	332	1176	.0904	3.54	30	9.50	.009516
		771	13016	1.0000				

$$A = \frac{\Sigma W_h\sqrt{C_h + 17M_h}}{\$110,000} = \frac{27.228792}{\$110,000} = .0002475$$

Final Allocation

(10) Institutions $\frac{W_h\sqrt{C_h + 17M_h}}{A}$ (n_h)	(11) Students n_hM_h	(12) Sampling Rate	(13) Approximate Sampling Fraction
28	3100	.777	4/5
33	1374	.471	1/2
45	943	.326	1/3
44	462	.226	2/9
38	135	.114	1/9
188	6014		

Table 2.7

Initial Sampling Design, and
Institutional Response as of January 14, 1965
to the First Mailing of Dec. 10, 1964

Stratum	Size	No. Invited	No. Replying	% Replying	No. Replying Favorably	% Favorable Replies
		PUBLIC INSTITUTIONS				
5	40+	20	14	70	14	100
4	20-39	14	9	64	8	89
3	10-19	15	15	100	14	93
2	5-9	13	12	92	9	75
1	0-4	9	9	100	6	67
	Total	71	59	83	51	86
		PRIVATE INSTITUTIONS				
5	40+	9	6	67	5	83
4	20-39	22	18	82	18	100
3	10-19	31	22	71	19	86
2	5-9	30	26	87	19	73
1	0-4	29	16	55	12	75
	Total	121	88	73	73	83
		TOTAL				
5	40+	29	20	69	19	95
4	20-39	36	27	75	26	96
3	10-19	46	37	80	33	96
2	5-9	43	38	88	28	74
1	0-4	38	25	66	18	72
	Total	192	147	77	124	84

Table 2.8

Worksheet for Second Sampling Allocation

(1) Stratum (h)	(2) "Size" of Institution	(3) No. of Institutions (from Table 2.4) (N_h)	(4) 142% of No. of Students in Table 2.4 (N_hM_h)	(5) Proportion of Students (W_h)	(6) Aver. No. of Students per Institution (M_h)	(7) Refigured Cost per Institution* (C_h)	(8) $\sqrt{C_h + 17M_h}$	(9) $\frac{W_h}{\sqrt{C_h + 17M_h}}$	Allocation (10) $\frac{\frac{W_h}{\sqrt{C_h + 17M_h}}}{A}$ (Institution) (n_h)	Allocation (11) Students (n_hM_h)	Allocation Sampling Rate
5	40+	36	3628	.3062	100.78	81	42.36	.007229	31	3124	.861
4	20-39	70	2653	.2239	37.90	44	26.24	.008533	36	1364	.514
3	10-19	138	2632	.2221	19.07	35	18.95	.011720	49	934	.355
2	5-9	195	1864	.1574	9.56	30	13.87	.011348	48	459	.246
1	0-4	332	1071	.0904	3.18	30	9.17	.009858	42	134	.127
		771	11848	1.0000					206	6015	

$$A = \frac{\Sigma W_h\sqrt{C_h + 17M_h}}{\$110{,}000} = \frac{26.066669}{\$110{,}000} = .0002369$$

Stratum (h)	Adjustment for Unfavorable Replies**	Adjustment for Non-Reply	Adjusted Sampling Rates	Adjusted No. of Institutions
5	1.05	1.30	1.175→1.000	36
4	1.04	1.30	.696	49
3	1.12	1.30	.514	71
2	1.35	1.30	.432	81***
1	1.38	1.30	.228	76
				313

*For convenience, these costs were taken from Table 2.6.

**Reciprocals of proportions of favorable replies, all institutions, by strata, Table 2.7.

***Through an error, this figure was originally computed as 64, and Table 2.9 reflects this error.

January 14. This figure was applied uniformly to strata since there seemed to be no consistent correlation between stratum and reply rate (see Table 2.7). (Through computational error, however, it was not applied to stratum 2; see Table 2.9 for further details.)

These final sampling rates were applied to the data of Table 2.4 in order to specify the numbers of institutions needed in a final, "expanded" sample as shown in Table 2.9 and the numbers of institutions that had to be added to the previous sample to achieve the expanded sample. In all, 101 institutions had to be added, according to these computations, and random sampling procedures were applied to select from a list of institutions for a "second wave" of invitations. Actually, mailings were made to 101 institutions on January 21, 1965, making a total of 293 institutions in the final sample.* As of 1962-63, there was a total of 5353 students who earned degrees in the five foreign languages; applying the growth ratio of 1.42 to this figure, we get a total of 7601 students in the estimated target population. It was not expected, however, that all

Table 2.9

Final Sampling Design

(Numbers of Institutions)

		PUBLIC				PRIVATE			
Stratum	Size	Pop.	Orig. Sample	Added	Expanded Sample	Pop.	Orig. Sample	Added	Expanded Sample
5	40+	25	20	5	25	11	9	2	11
4	20-39	27	14	5	19	43	22	8	30
3	10-19	45	15	9	24	93	31	15	46
2	5-9	61	13	7	20*	134	30	13	43*
1	0-4	75	9	8	17	257	29	29	58
		233	71	34	105	538	121	67	188

TOTAL

Stratum	Size	Pop.	Orig. Sample	Added	Expanded Sample
5	40+	36	29	7	36
4	20-39	70	36	13	49
3	10-19	138	46	24	70
2	5-9	195	43	20	63*
1	0-4	332	38	37	75
		771	192	101	293

*These are the actual figures used. It was later discovered, however, that a computational error had been made. The correct figures for the expanded sample should have been: Public, 25; Private, 56; Total, 81.

---

*Although it did not happen to appear in the random sample drawn, Harvard and Radcliffe Colleges were also participating institutions. As had been true the previous year, they contributed very few students to the sample due to student non-response.

these students would be tested because of both institutional and student non-response rates. Chapter III presents and discusses the outcomes of the sampling design.

4. Procedures in Data Collection

The success of this study depended in large part on the excellent cooperation of many individuals at numerous institutions of higher learning in the United States. They were mainly responsible for the actual collection of the data. A great deal of administrative work, however, had to be done at the project headquarters in order to insure a smoothly running program. Too, a considerable amount of work was done by personnel at Educational Testing Service in shipping test materials and scoring the test results. This section describes the sequence of events planned and carried out by all parties concerned.

(1) Once an institution was selected for participation in the study, the project headquarters at Harvard mailed to the president or chief executive of that institution the following items: (these items, except for reply envelopes and the like, are included in this report in Appendix D)

a. Five copies of a covering Memorandum to Presidents of Institutions of Higher Learning signed by Dr. Kenneth Mildenberger, then Director of the Division of College and University Assistance, U. S. Office of Education, pointing out the need for and the importance of the project and encouraging the institution to participate; it requested each President to appoint a coordinator for the testing project at his institution;

b. Five copies of a Memorandum Describing the Project, that could be distributed to foreign language department chairmen and others who might be concerned with the decision whether or not the institution would participate in the project;

c. An Institution Information Form asking whether the institution would participate in the project, in what languages the institution offered "bachelor's degree programs" and how many seniors were currently "majoring" in each language; if the institution had decided to participate in the testing program, the form also asked for the name and mailing address of the person who had been designated as coordinator for the program and for the names and addresses of each FL department chairman. An early deadline was given for the return of the form.

d. A business reply envelope for return of the Institution Information Form.

Comment: The decision to mail materials directly to the chief executives of the institutions concerned was based partly on the fact that no comprehensive list of foreign language department chairmen was available. Furthermore, it was reasoned that presidents should have information about the project and have the right to make the final decision as to the participation of their institutions. In the New York State pilot study, much more use of direct communication with chairmen of foreign language departments was employed, with perhaps slightly better results. One faces a fundamental dilemma in a study of this type: while the chief executives of institutions are the proper authorities to address, for many reasons, they are busy people and frequently have poor communication with their own department chairmen, who in many cases would have more interest in such a study than the presidents.

As noted earlier, the first wave of mailings was made to 192 institutions on December 10, 1964, and the second wave was made to 101 institutions on January 21, 1965, in anticipation of testing dates of institutions that would occur as early as possible in the spring of 1965. It is impossible to say whether higher institutional reaponse rates would have occurred if it had been possible to make the mailings earlier.

(2) On receipt of the above kit of materials and after deliberation as to participation in the study, the institution returned the completed Institution Information Form to project headquarters at Harvard.

Comment: The fact that this form was to be sent directly to Harvard rather than to ETS reflected the decision to centralize as much as possible of the administrative work at Harvard in order to make coordination more effective and to minimize administrative charges.

(3) Beginning January 15, 1965, upon receipt of the Institution Information Form from institutions that had decided to participate, the project headquarters mailed to the designated institutional Coordinator the following kit of materials:

a. (Mimeographed) "Instructions and Suggestions to Coordinator for Administering Test Materials under the College Foreign Language Testing Project, Spring 1965" --a "master information source for all activities and responsibilities of the institution coordinator" (Appendix D-4);

b. One copy of "Order Form for Administration of MLA Foreign Language Proficiency Tests for Teachers and Advanced Students under the College Foreign Language Testing Project"--to be returned to project headquarters at Harvard in an airmail business reply envelope that was supplied (Appendix D-5);

c. One "Application for Reimbursement of Coordinators and Assisting Proctors"--to be returned to project headquarters after the administration of the tests;

d. A sufficient supply of the booklet "A Description of the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students (ETS, 1964a) to distribute to each foreign language major graduating from the institution in the spring of 1965 in one of the five test languages;

e. An equal number of copies of the leaflet "A Letter to College Seniors Graduating in the Spring of 1965 with Degrees in French, Italian, Spanish, German, or Russian"--also to be distributed to students (Appendix D-6);

f. An envelope containing materials to be distributed by the Coordinator to foreign language department chairmen; it contained for each department chairman (1) a copy of the above-mentioned "Memorandum Describing the Project" (Appendix D-2), (2) the "Questionnaire for Chairmen of Foreign Language Departments" (Appendix C), and (3) a business reply envelope for the use of the chairmen in returning the questionnaire to project headquarters.

Comment: The Coordinator had the responsibility of arranging the testing dates, distributing materials to department chairmen, informing students about the project or seeing that they were informed, ordering test supplies, arranging for the actual administration of the tests, and returning materials either to ETS (the completed tests materials) or to Harvard (the student questionnaires and the vouchers for payment of himself and proctors). For this he was paid $30; at the larger institutions this was probably hardly adequate to recompense the coordinators for the amount of time they had to devote to the project. Coordinators at the larger institutions were permitted to employ assisting proctors, to be paid $10 apiece.

It was specified that testing dates were to fall within the period March 1 to May 8, 1965 and were to be set in such a way as to allow sufficient time for testing materials to be shipped to the institution from ETS.

The Coordinator was conceived to be the key factor in encouraging students to participate in the testing. It was recommended to Coordinators that they try to contact students personally. There is no information available as to the extent to which they actually did so. As will be seen, student response varied widely from institution to institution.

(4) Upon receipt of the test order form, the project headquarters arranged with ETS to send to the Coordinator a sufficient supply of MLA Proficiency Tests (in the relevant languages), Modern Language Aptitude Tests, and Student Questionnaires to cover the numbers of students expected to be tested. (A small excess supply was provided in accordance with ETS experience in this matter.) The packages of test materials included directions for administering the tests. Sent directly from Harvard to each institution was a copy of Directions for Administering the Questionnaire for Foreign Language Majors (Appendix D-7).

(5) The actual distribution of weeks in which testing dates selected by institutions fell was as follows:

March 1-6	1
March 8-13	5
March 15-20	13
March 22-27	19
March 29-April 3	31
April 5-10	50
April 12-17	17
April 19-24	34
April 26-May 2	27
May 3-8	12
May 10-15	2
	211

The excess total number in the above table (as compared to the 202 institutions in the study) reflects the fact that a few institutions scheduled different parts of the testing program in two different weeks.

Normally, all students who volunteered to participate in the study took the four "skills" tests of the MLA Proficiency Series in the language they claimed as their major language, the "short form" of the Modern Language Aptitude Test, and the Student Questionnaire. There were no specifications as to the order in which these tests were administered except that the Listening and Speaking tests of the MLA series were to be given always in that order. Students enrolled in teacher preparation programs also took the three professional knowledge tests of the MLA series.

(6) Upon termination of the testing, the MLA and MLAT materials--both used and unused--were returned to ETS, where the answer sheets were scored, and the Student Questionnaires were returned directly to Harvard, where the information contained therein was key-punched.

(7) ETS supplied Harvard with cards containing the scores for the tests. Because of the considerable amount of time required to score the Speaking tests (which required students to record their responses on tape), ETS was not able to supply the bulk of the scores until September, 1965. However, special arrangements had been made to score the tests of students at institutions in Pennsylvania as early as June so that the scores could be supplied to department chairmen at those institutions for use in the state of Pennsylvania teacher certification program. The scores supplied by ETS were "converted scores" on the MLA Proficiency Tests, but raw scores on the three parts of the MLAT that were administered.

(8) Using the addresses that the students had supplied on the Student Questionnaire (Appendix B), the project personnel made up individual score reports and mailed them to the students on October 7, 1965. The score reports were in the form of pressure-sensitive labels, imprinted by tabulating equipment, that were affixed to a Memorandum to the Foreign Language Majors who Took the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students in the Spring of 1965 (see sample, Appendix E-1). Each score report was accompanied by an interpretative leaflet to enable the student to assess his performance in comparison to norms compiled for NDEA Institute students (Appendix E-2). Scores on the Modern Language Aptitude Test were not furnished to students; it was explained that these scores were collected only for statistical control purposes.

(9) Rosters of students' MLA test scores were made up at the project headquarters and mailed to foreign language department chairmen on October 22, 1965 with a covering letter (Appendix E-3) and interpretative leaflet (Appendix E-2).

(10) Department Chairmen's Questionnaires were returned directly to Harvard by the department chairmen themselves.

(11) Data collected were further processed for analysis at project headquarters. Various statistical tabulations were made over a period of at least a year because of the large volume of data and the many cross-comparisons and correlations that were desired.

(12) The data have been archived on magnetic tape and are deposited with the ERIC Clearinghouse on Foreign Languages, Modern Language Association, 62 5th Ave., New York 10011, N.Y. for use by qualified investigators. The tapes are accompanied by appropriate documentation as to card codes and formats.

Throughout the above steps, it was frequently necessary to send reminder letters to the various key persons involved in order to secure the best possible response.

Chapter III

OUTCOMES OF THE SAMPLING DESIGN

This chapter is concerned solely with general matters pertaining to the amount and character of data accumulated as a result of the sampling design and the procedures for gathering data. It does not attempt to estimate the amount of bias in the sample or correct for it; this is done in Chapter V.

1. Institutional response rates

According to Dr. Peter Loret of Educational Testing Service, who participated in a June 1966 conference to inspect the preliminary results of this study, ETS had found that the institutional response rate could be regarded as a more or less permanent "institutional characteristic." Some institutions are consistently willing to participate in educational testing surveys, while others are consistently unwilling. In his experience, an overall response rate of 50% was typical.

In the present study, 203 of the 293 institutions invited to participate did so; the institutional response rate was thus 69.3%. Detailed figures are presented, by strata and the public-private dichotomy, in Table 3.1. Although there was little difference between public and private institutions in rate of response, there was a clear trend whereby the larger institutions were more likely to agree to participate.

Comparison with the response rates obtained in the New York pilot study of 1964 are of interest (Carroll, 1965). The invitations to institutions were mailed out much later in the year in that study--on March 24, 1964, and the testing dates were typically in early May. Still, the overall institutional response rate was quite good--70.1%. It had been concluded that it might have been better if the institutions had been invited much earlier in the year, and it was on this basis that the present study began inviting institutions early. Nevertheless, the national institutional response rate was no better than that of the New York pilot study. A possible explanation for this is the keen interest of New York institutions in the MLA Tests in connection with teacher certification requirements, as compared with the general level of interest in the MLA tests throughout the country. (In the national study, the response rate of New York institutions was 77.4%.)

Response of institutions to the mailings was fairly prompt. All the favorable replies were received within eleven weeks of the initial mailing. Replies to the first wave of invitations were somewhat prompter than those to the second, and also, a higher proportion of favorable replies were received to the first wave of invitations (146 out of 192, or 76.0%) than to the second (60 out of 101, or 59.4%). Table 3.2 gives the details.

Five institutions initially agreed to participate, but later found it necessary to withdraw. In the tabulations presented here, they are treated as non-participants. As of April 9, 1965, only 10 institutions had not given a reply--even after reminder letters and telephone calls (1 from the first sample, 9 from the second sample). Further efforts to get these institutions to participate were abandoned since it was by then too late to arrange for testing.

Table 3.1

Institutional Response Rates

			SOLICITED FINAL SAMPLE		PARTICIPATING		
Stratum	Size*	No. in Population	N	% of Population (Actual Sampling Rates)	N	% of Sample (Response Rate)	% of Population (Net Sampling Rate)
				PUBLIC			
5	40+	25	25	100.0	20	80.0	80.0
4	20-39	27	19	70.4	13	68.4	48.1
3	10-19	45	24	53.3	17	70.8	37.8
2	5-9	61	20	32.8	11	55.0	18.0
1	0-4	75	17	22.7	11	64.7	14.7
Total		233	105	45.1	72	68.6	30.9
				PRIVATE			
5	40+	11	11	100.0	10	90.9	90.9
4	20-39	43	30	69.7	26	86.7	60.5
3	10-19	93	46	49.5	32	69.6	34.4
2	5-9	134	43	32.1	30	70.0	22.4
1	0-4	257	58	22.6	33	56.9	12.8
Total		538	188	34.9	131	69.7	24.3
				TOTAL			
5	40+	36	36	100.0	30	83.3	83.3
4	20-39	70	49	70.0	39	79.6	55.7
3	10-19	138	70	50.7	49	70.0	35.5
2	5-9	195	63	32.3	41	65.1	21.0
1	0-4	332	75	22.6	44	58.7	13.3
Total		771	293	38.0	203	69.3	26.3

*Number of FL majors in all five languages as of 1962-63.

Table 3.2

Rate of Return of Favorable Replies from Institutions

	First wave of invitations (192 Mailed Dec. 10, 1964)			Second wave of invitations (101 Mailed Jan. 21, 1965)		
Wks. after mailing	Wk. ending	No. of favorable replies	Cumulative per cent	Wk. ending	No. of favorable replies	Cumulative per cent
1	Dec. 18	7	3.6	Jan. 29	0	0.0
2	Dec. 25	47	28.1	Feb. 5	17	16.8
3	Jan. 1	26	41.7	Feb. 12	19	35.6
4	Jan. 8	13	48.4	Feb. 19	14	49.5
5	Jan. 15	23	60.4	Feb. 26	4	53.5
6	Jan. 22	2	61.5	Mar. 5	1	54.4
7	Jan. 29	0	61.5	Mar. 12	1	55.4
8	Feb. 5	5	64.1	Mar. 19	3	58.4
9	Feb. 12	16	72.4	Mar. 26	1	59.4
10	Feb. 19	5	75.0			
11	Feb. 26	2	76.0			
	Total	146	76.0		60	59.4

2. Student response rates

It will be recalled that the design of the sample called for a two-stage sampling -- a stratified optimal-allocation sample of institutions at the first stage and a 100% sample of students at the second stage. Every effort was made to encourage all graduating FL majors at the selected participating institutions to take the tests, but despite these efforts, the student response rates were far short of what was desired.

Evidence as to the numbers of students available for testing at the institutions was available from two sources: (1) figures provided by the departments themselves in connection with the administration of the tests (see Item 2, Institution Information Form, Appendix D-3), and (2) official figures on earned degrees at U. S. Colleges and universities collected by the U. S. Office of Education and generally published one or two years after the graduation date (for example, see Tolliver, 1963a, 1963b; Wright, 1965, 1966). As of this writing, the official figures for the academic year 1964-65 have not been published[1], but tabulator cards containing the relevant information were supplied to the project in July, 1966. In view of the fact that these latter figures are of an official character and cover both participating and non-participating institutions, it was decided to rely on these, rather than those supplied by department chairmen, to make an analysis of student response rates. (Reliance was placed in the accuracy of the data in the cards; it may eventually be found that there are slight discrepancies with published figures.)

In making this analysis, however, it was necessary to exclude students at institutions that had not been included in the 1962-63 population of institutions from which the sample had been drawn, since only in this way could the success of the sampling design be fairly judged. The U. S. Office of Education figures showed a total of 12,220 students as having earned B. A. degrees in the academic year 1964-65, but only 11,633 of these were in institutions defined as being in the 1962-63 population of institutions.

The analysis of student response rates had to take account also of the fact that an examination of the total CFLT sample utilized in various tabulations later in this report included 3 students tested at Harvard University, an institution that had not been drawn in the random sample of institutions but where testing was offered as a matter of courtesy. (The 3 students represented, actually, only a small fraction of those available to be tested at the institution.) The final count of graduating senior students legitimately tested at institutions that had been drawn in the sampling design was 2784.

From the U. S. Office of Education figures, it was determined that there were 5427 students available to be tested in 1965 at the institutions participating in the study. The overall student response rate, therefore, was 51.3%. The tested students represented, in all, 23.9% of the 11,633 students in the target population of French, German, Italian, Russian, and Spanish majors in the 1962-63 population of institutions.

Table 3.3 shows a detailed breakdown of the student response rates by size-stratum, institution type, and sex. Several sets of percentages may be considered. First we may consider the percentages that the FL majors in the participating institutions constituted of the total target population as defined above. These percentages reflect the sampling design whereby larger sampling fractions were used for the drawing of the larger institutions. They also reflect the institutional response rates discussed in the previous section. If all students at the participating institutions had volunteered to take the tests, as the sampling design had contemplated, these percentages would have reflected the degree to which each stratum of the target population was represented in the tested sample. They would have deviated from the optimal sampling design only to the extent that institutional response rates varied. It can be seen that the "available student" sub-population was somewhat more representative of students at public institu-

---

[1] The citation to the publication, in press at this writing has been furnished by the U. S. Office of Education. See Mason and Rice (1967), in the References.

Table 3.3

Student Response Rates, by Strata, Institution Type, and Sex

Stratum	Size as of '62-'63	Total Population of FL Majors (1964-65)[1]			FL Majors in Participating Institutions						Tested Students								
					N's			Percent of Pop.			N's			% of Sample			% of Pop.		
		M	F	Tot.	M	F	Tot.	M	F	Tot.	M	F	Tot.	M	F	Tot.	M	F	Tot.
								PUBLIC INSTITUTIONS											
5	40+	604	1993	2597	453	1565	2018	75.0	78.5	77.7	129	480	609	28.5	30.7	30.2	21.4	24.1	23.5
4	20-39	310	754	1064	170	336	506	54.8	44.6	46.7	65	169	234	38.2	50.3	46.2	21.0	22.4	22.0
3	10-19	257	756	1013	88	268	356	34.2	35.5	35.1	59	139	198	67.0	51.9	55.6	23.0	18.4	19.5
2	5-9	259	530	789	55	144	199	21.2	27.2	25.2	36	109	145	65.5	75.7	72.9	13.9	20.6	18.4
1	0-4	106	249	355	28	54	82	26.4	21.7	23.1	15	52	67	53.6	96.3	81.7	14.2	20.9	18.9
TOTAL		1536	4282	5818	794	2367	3161	51.7	55.3	54.3	304	949	1253	38.3	40.1	39.6	19.8	22.2	21.5
								PRIVATE INSTITUTIONS											
5	40+	216	526	742	208	492	700	96.3	93.5	94.3	82	230	312	39.4	46.7	44.6	38.0	43.7	42.0
4	20-39	363	874	1237	130	464	594	35.8	53.1	48.0	92	315	407	70.8	67.9	68.5	25.3	36.0	32.9
3	10-19	300	1182	1482	129	423	552	43.0	35.8	37.2	84	326	410	65.1	77.1	74.3	28.0	27.6	27.7
2	5-9	360	921	1281	59	197	256	16.4	21.4	20.0	58	184	242	98.3	93.4	94.5	16.1	20.0	18.9
1	0-4	313	760	1073	61	103	164	19.5	13.6	15.3	61	99	160	100.0	96.1	97.6	19.5	13.0	14.9
TOTAL		1552	4263	5815	587	1679	2266	37.8	39.4	39.0	377	1154	1531	64.2	68.7	67.6	24.3	27.1	26.3
								TOTAL											
5	40+	820	2519	3339	661	2057	2718	80.6	81.7	81.4	211	710	921	31.9	34.5	33.9	25.7	28.2	27.6
4	20-39	673	1628	2301	300	800	1100	44.6	49.1	47.8	157	484	641	52.3	60.5	58.3	23.3	29.7	27.9
3	10-19	557	1938	2495	217	691	908	39.0	35.7	36.4	143	465	608	65.9	67.3	67.0	25.7	24.0	24.4
2	5-9	619	1451	2070	114	341	455	18.4	23.5	22.0	94	293	387	82.5	85.9	85.1	15.2	20.2	18.7
1	0-4	419	1009	1428	89	157	246	21.2	15.6	17.2	76	151	227	85.4	96.2	92.3	18.1	15.0	15.9
TOTAL		3088	8545	11633[2]	1381	4046	5427	44.7	47.3	46.7	681	2103	2784[3]	49.3	52.0	51.3	22.1	24.6	24.0

---

[1] This population is restricted to institutions existing in the 1962-63 population used for the sampling design. It is derived from U. S. Office of Education figures, but includes in addition the students at four institutions in the 1962-63 samples that participated in the national study but that were not included in the U. S. Office of Education data. Data on the numbers of these students came from information supplied by the department chairmen at those institutions.

[2] In addition there were 60 students listed as majoring in Romance Languages.

[3] Does not include 3 students tested at Harvard, which was not one of the sampled institutions.

tions (39.6%); this difference is ascribable to differences between public and private institutions in the distribution of students over strata of different sizes. The sample would have been about equally representative of male and female foreign language majors, despite the proponderance of females in the target population.

Of most interest is the next set of percentages, showing the actual student response rates in institutions of different sizes and types. Three major conclusions emerge from these figures:

(1) There was no appreciable or consistent difference in the response rates of men as compared to women. (See Technical Note below.)

(2) There was a strong negative relationship between response rate and institution size: this was true in both public and private institutions. That is, student response rates were relatively low in the larger institutions and relatively high in the smaller institutions.

(3) The response rates were higher at private (67.3%) than at public institutions (39.6%), and the difference was consistent from stratum to stratum. This finding possibly suggests that student-faculty relations at private institutions were more propitious for establishing the personal contacts that would lead to student volunteering to take the tests in the study.

> Technical note: The differences in reponse rates by sex, institution-type and stratum were examinable for statistical significance by the chi-square test. In the case of the sex variable, the value of chi-square for the total sample was 2.65, with 1 d.f., not significant at the 5% level even though there was a slight difference favoring the women. (The chi-square test was also applied to sex differences in response rates within particular strata and institution-type classifications. Significant differences favoring the women at the 5% level or better were found in Public Stratum 4, Public Stratum 1, and Private Stratum 3, while a significant difference favoring the men was found in Public Stratum 3. These findings do not add up to any consistent pattern and may therefore be regarded as reflecting variations of no practical interest.) The chi-square values for the public-private dichotomy and the institution-size classification were both very highly significant, 404.2 with 1 degree of freedom, and 806.0 with 4 degrees of freedom, respectively.

The large differences in student response rates between institution types and sizes were, of course, a matter of considerable concern with respect to the validity of the sample, but there was nothing that could reasonably have been done post facto to alter the sample. Reliance therefore had to be placed on statistical analyses of the data that were at hand in order to estimate how much student response-rate differences might have biased the sample. This problem is taken up in Chapter V.

The final set of percentages listed in Table 3.3, in the last three columns, show what parts of the total population, by sex, institution-type, and stratum, were obtained in the CFLT sample. The last figure in the table, for example, is 24.0%, indicating that 24% of the 11,633 FL majors in the target population appeared in the CFLT sample. This is not to say, of course, that the 24% appearing in the CFLT sample were necessarily representative of the total population, although the sample design had been developed with the intention of drawing a representative sample. A further intention of the sampling design, however, had been to drawn a sample constituting about 50% of the target population. This aim was achieved only about half-way; it was frustrated by the student response-rates that were lower than desired.

Actually, the combination of sampling design sampling fractions that increased with institution size and student response rates that decreased with institution size resulted in a series of proportions that were fairly constant over strata and institution types.

The net effect is that the CFLT sample has the appearance of being sampled from stratum, institution-type, and sex classifications with approximately equal sampling fractions (ranging only from 15.9% to 27.9% for the 5 strata when institution-types are combined). If (and this is a "big if") student response rate is uncorrelated with the dependent variables of the study, the CFLT sample can be used as it stands with no correction for student response-rates or sampling fractions in the sampling design, as reasonably representative of the target population. As will be shown in Chapter V, it is unlikely that student response rate is highly associated with the chief dependent variables of this study (the MLA skills tests), although there is probably a small degree of association. For this reason we feel that there is probably little error introduced by using the CFLT sample as it stands, for many of the analyses and interpretations made in this study(as is done in all chapters except Chapter V).

Table 3.4 shows an analysis of student response by language. A chi-square test shows that there is a highly significant (p <.001) difference in these response rates. The difference, however, appears to be largely associated with Russian, where the response rate was consistently low in both public and private institutions. No suggested explanation can be offered for this finding, nor can we suggest what implications it has as to possible special biases of the Russian sample.

3. Completeness of Data and Preparation of the Sample

In a study of this type, conducted with the cooperation of many different persons at institutions spread throughout the U. S., it is inevitable that the data are not as complete as might be desired, even for those students who actually participated in the study. The testing program was long, and it was probably difficult for institutional coordinators to insure that all participating students took every test and filled out a questionnaire. "Missing data" pose a problem for the statistical analyst. For some types of questions, he attempts to estimate what the data would show if none were missing. For other types of questions, it suffices to analyze the data at hand and assume that they are representative of what might have been obtained if the data were complete.

There were three basic instruments for students in the present study: the MLA Proficiency Test in the relevant language, the Modern Language Aptitude Test, and the Student Questionnaire. When all the data were all assembled, it was found that there were 2880 different individuals who contributed data on some or all of these instruments. There were distributed as follows:

	N	%
French	1300	45.1
German	438	15.2
Italian	31	1.1
Russian	119	4.1
Spanish	992	34.4
	2880	100.0

Examination of these data, however, showed that 58 (2.0%) were underclassmen and 35 (1.2%) had already graduated and therefore should not have been tested. These cases were eliminated, leaving 2787 legitimate cases. All these cases had student questionnaire data, except that for a few of these, there were missing entries in the questionnaires.

Of the 2787 legitimate cases:

2587 (92.8%) had scores on all four skills tests (Parts 1-4)

193 (6.9%) had some scores missing on Parts 1-4 (most of the missing scores being for the Speaking Test because of poorly recorded tapes)

7 had no scores on MLA skills tests (of these, 2 had scores on the teacher preparation tests)

Table 3.4

Student Response Rates, by Language

Language	No. in Population ('64-'65)	No. in Institutions of Final Sample N	No. in Institutions of Final Sample %	Tested Students N	Tested Students % of Sample	Tested Students % of Population
			PUBLIC			
French	2231	1171	52.5	496	42.4	22.2
German	930	535	57.5	188	35.1	20.2
Italian	53	40	75.5	10	25.0	18.9
Russian	290	206	71.0	55	26.7	19.0
Spanish	2314	1209	52.2	504	41.7	21.8
TOTAL	5818	3161	54.3	1253	39.6	21.5
			PRIVATE			
French	2812	1116	39.7	774	69.4	27.5
German	818	310	37.9	225	72.6	27.5
Italian	55	24	43.6	18	75.0	32.7
Russian	266	125	47.0	50	40.0	18.8
Spanish	1864	691	37.1	464	67.1	24.9
TOTAL	5815	2266	39.0	1531	67.6	26.3
			TOTAL			
French	5043	2287	45.3	1270	55.5	25.2
German	1748	845	48.3	413	48.9	23.6
Italian	108	64	59.3	28	43.8	25.9
Russian	556	331	59.5	105	31.7	18.9
Spanish	4178	1900	45.5	968	50.9	23.2
TOTAL	11,633*	5427	46.7	2784	51.4	23.9

*In addition, there were 60 students listed as majoring in Romance Languages. Some of these may have appeared in the tested sample, but information is not available as to which language they were tested in.

Further, of the 2787 legitimate cases:

1116 (40.0%) took the teacher preparation sections of the MLA tests (parts 5-7) presumably because they were enrolled in teacher preparation programs (there were a handful of missing scores):

1671 (60.0%) did not take any of the teacher preparation tests, presumably because they were not enrolled in teacher preparation programs.

2534 (90.9%) of the 2787 students took the MLAT, there being only 5 cases with some missing scores.

In all, there were 2389 (85.7%) complete on MLA skills test scores and the MLAT; of these, 980 were also complete on the teacher preparation sections of the MLA battery.

An examination of the Student Questionnaires showed that there were a number of individuals in the total sample who might not be regarded as "regular" cases who were native speakers of English and who took their beginning and intermediate foreign language instruction in regular schools. All 2880 cases were divided into a number of types: three "regular" types and six "odd" types, as follows:

"Regular" types

1	Student began the study of his major FL in "grade school"	282	10.8%
2	Student began the study of his major FL in secondary school	1662	63.8%
3	Student began the study of his major FL in college	660	25.3%
	Total . . .	2604 (90.4%)	99.9%

"Odd" types

4	The first course the student listed as having taken in his major FL was of an advanced type, from which fact it might be presumed that he had a non-usual introduction to the language	101
5	The student's experience included intensive training in his FL in military service	13
6	The student's experience included two or more years spent in a country where the FL is spoken . . .	17
7	Course data were missing or sparse	66
8	Miscellaneous unusual cases	21
9	Native speakers of the FL	58
	Total	276 (9.6%)

Assignment of a student to an "odd" type category was made if there was any evidence to support this even though he might have listed beginning courses that would otherwise have led to his assignment to "regular" type categories. This was done in order to remove any doubt that the "regular" cases might contain cases with irregular or unusual experiences in the relevant foreign language. Most of the analyses to be reported here are based only on "regular" cases, or upon subgroups of these.

Chapter IV

RESULTS FROM TESTS OF FOREIGN LANGUAGE ATTAINMENT:

THE TESTED SAMPLE

This chapter will concern only the results of the four MLA skills tests actually given to the sample that volunteered to take these tests. It will show the overall statistics, frequency distributions, and intertest correlations obtained for the "regular" and "odd" cases, without consideration of the fact that the tested sample was undoubtedly biased in several ways with respect to the target national population. Chapter VI will present attempted corrections for these sampling biases. Furthermore, this chapter will give scant attention to various breakdowns of the sample that could be made, since these will be the concern of later chapters that investigate the association of the test scores with various student background and instructional variables.

Since the MLA skills tests constitute the chief criteria used for evaluating the foreign language attainments of the sample, it is well to focus attention on the general characteristics of the test results. Can we tell, from these characteristics, anything about how well the tests appear to measure foreign language attainment? Can we say anything about whether all four tests--Listening, Speaking, Reading, and Writing-- measure the same or different aspects of achievement? These matters will be taken up in this chapter.

1. Distributional statistics, all cases

Table 4.1 shows distributional statistics for the MLA skills test scores for all cases in the sample ("regular" and "odd") that had complete data on these four tests, for 1170 cases in French, 388 cases in German, 102 cases in Russian, and 900 cases in Spanish. No data are shown for Italian because of the small number of cases. Figures 4.1, 4.2, 4.3, and 4.4 depict the score distributions graphically: the scores are plotted in accordance with the MLA-FSI equivalents established in Chapter II, in order to facilitate comparisons across skills and across languages.

The scores range from values that are very low--lower than the estimated "chance" scores for tests where such scores can be estimated--to values that are at or near the maximum possible scores. It is evident, however, that some of the tests have a "ceiling," at least in terms of the FSI equivalents, whereby they cannot measure differences among the upper FSI ratings. This is universally the case for the Listening test, which measures only up to S-3+ in French and Russian, S-4 in Spanish, and S-4+ in German. Speaking measures only up to S-4+ in German and Russian; Reading fails to cover R-5 in French and Russian, and Writing fails to cover R-5 in Spanish and the range from R-3+ to R-5 in Russian. These ceilings are also reflected to a considerable extent in the negatively skewed distributions for some of these tests. For example, the distribution of scores for French Listening seems to be sharply cut off at its maximum score, which corresponds to an FSI rating of S-3+. If the test had been constructed with more difficult items it might have been able to discriminate among the upper FSI ratings. Indeed, all the distributions of Listening test scores, save that in Russian, are negatively skewed to a highly significant extent, and so are most of the distributions of the Writing test scores. Although the MLA tests are billed as "advanced" tests, in many cases they are not sufficiently difficult to discriminate among really advanced levels of proficiency. This fact must be borne in mind as the reader examines data to be reported here. Furthermore, the negative skewnesses of many of the test score distributions may make means and standard deviations somewhat misleading. In a great many of the statistical analyses to be reported here, means and standard deviations are employed because they are statistics reported by the available computer programs; medians and interquartile ranges would perhaps be more appropriate statistics to report in the case of the negatively skewed distributions.

Table 4.1

Data from MLA Skills Test Score Distributions
(All Cases Complete on 4 MLA Skills Tests)

Statistic	Listening	Speaking	Reading	Writing
FRENCH (N = 1170)				
Max. Possible Score (Form A)	56	125	70	80
Highest Obtained Score	56	115	69	74
Median (Estimated)	45.3995	73.6550	49.3089	47.3355
Mean	44.3051	73.9248	48.9991	47.3949
Lowest Obtained Score	24	39	27	22
Chance Score	29	--	32.5	--
Standard Deviation	7.6219	9.5323	8.6354	9.6519
g_1 (Measure of Skewness)	-0.4308**	0.0849	-0.1076	0.0184
GERMAN (N = 388)				
Max. Possible Score (Form A)	56	125	70	80
Highest Obtained Score	56	124	70	77
Median (Estimated)	46.3129	90.2181	53.7248	56.0342
Mean	44.7990	89.7552	52.6753	54.3144
Lowest Obtained Score	24	45	25	20
Chance Score	29	--	32.5	--
Standard Deviation	8.6787	14.4050	10.0879	13.5483
g_1 (Measure of Skewness)	-0.5233**	-0.0964	-0.3121*	-0.3808**
RUSSIAN (N = 102)				
Max. Possible Score (Form A)	56	125	70	80
Highest Obtained Score	56	122	70	80
Median (Estimated)	43.1607	77.8631	41.5714	65.9171
Mean	43.2843	80.8431	43.3431	63.3922
Lowest Obtained Score	32	50	24	35
Chance Score	29	--	32.5	--
Standard Deviation	6.2141	14.7534	11.1536	11.2050
g_1 (Measure of Skewness)	0.0597	0.6060*	0.4765*	-0.6760**
SPANISH (N = 900)				
Max. Possible Score (Form A)	56	125	70	80
Highest Obtained Score	56	122	69	78
Median (Estimated)	45.9926	85.1313	48.3385	56.2052
Mean	44.9722	85.1800	48.6044	55.4556
Lowest Obtained Score	24	46	27	24
Chance Score	29	--	32.5	--
Standard Deviation	6.8664	12.7835	8.7697	10.5109
g_1 (Measure of Skewness)	-0.4458**	0.0114	0.0910	-0.2140*

**p < .01, *p < .05 for hypothesis of $g_1 = 0$ and normality of distribution.

Figure 4.1

Frequency Distribution of Scores on MLA Skills Tests, French, Tested Sample

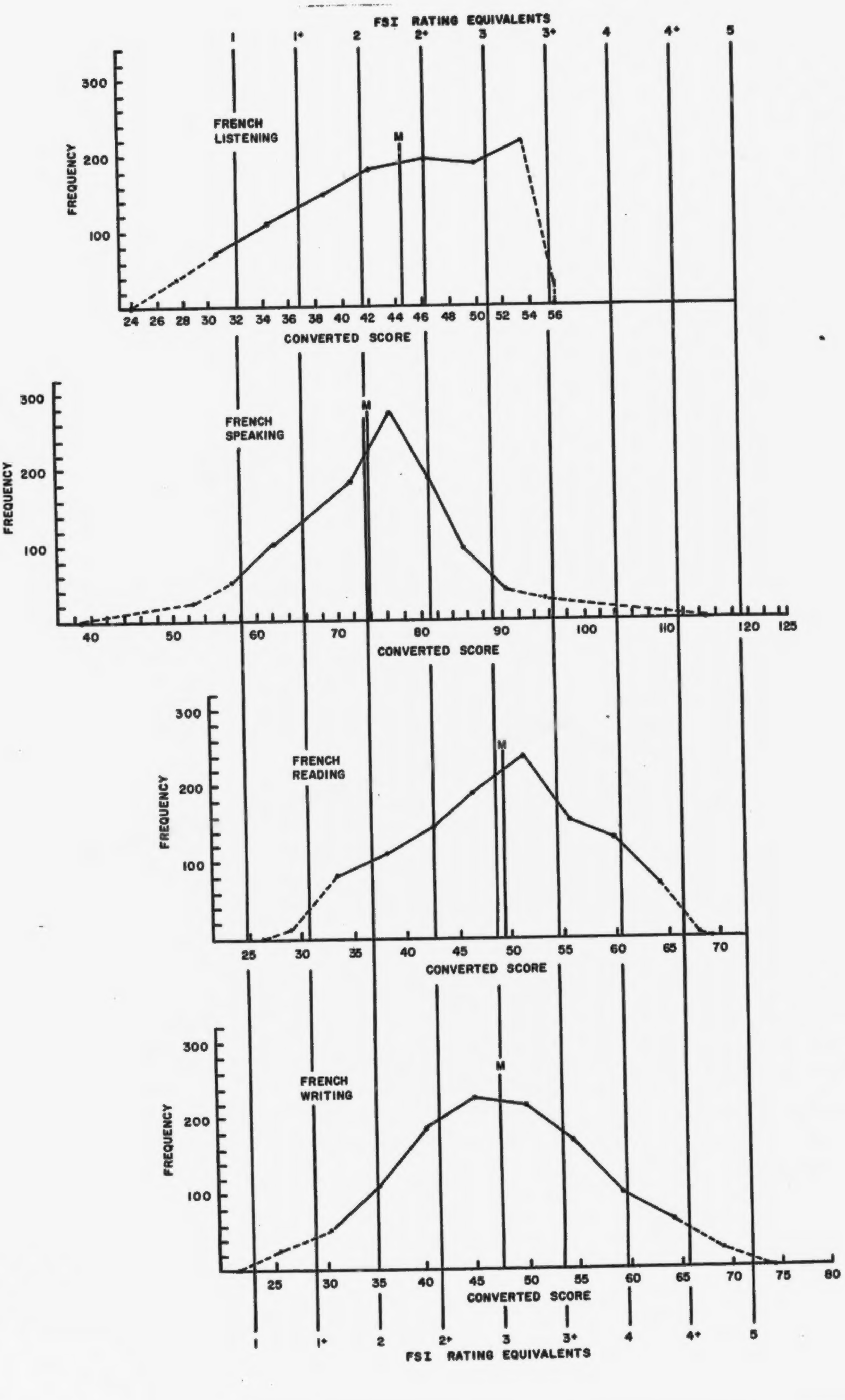


Figure 4.2

Frequency Distribution of Scores on MLA Skills Tests, German, Tested Sample

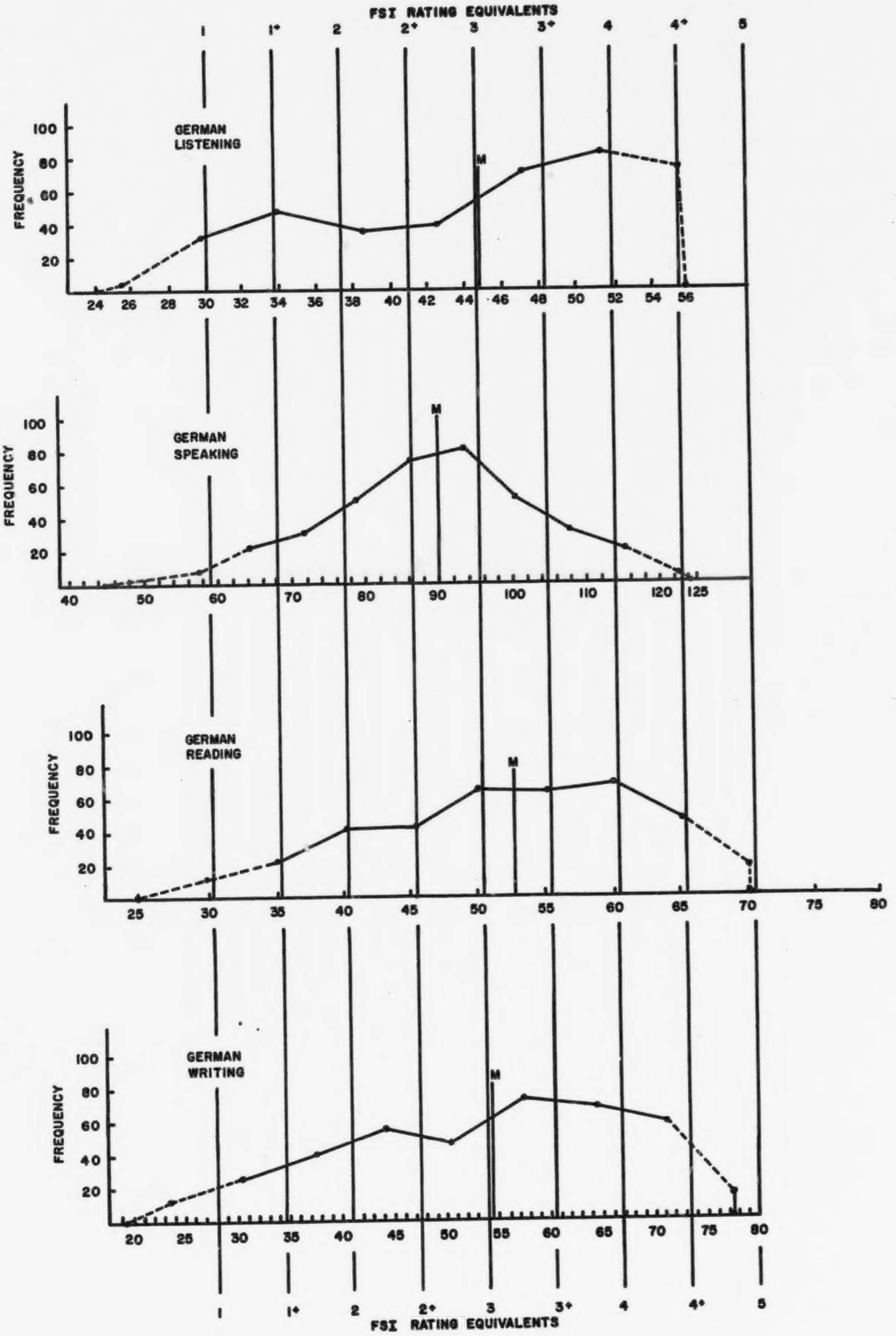


Figure 4.3

Frequency Distribution of Scores on MLA Skills
Tests, Russian, Tested Sample

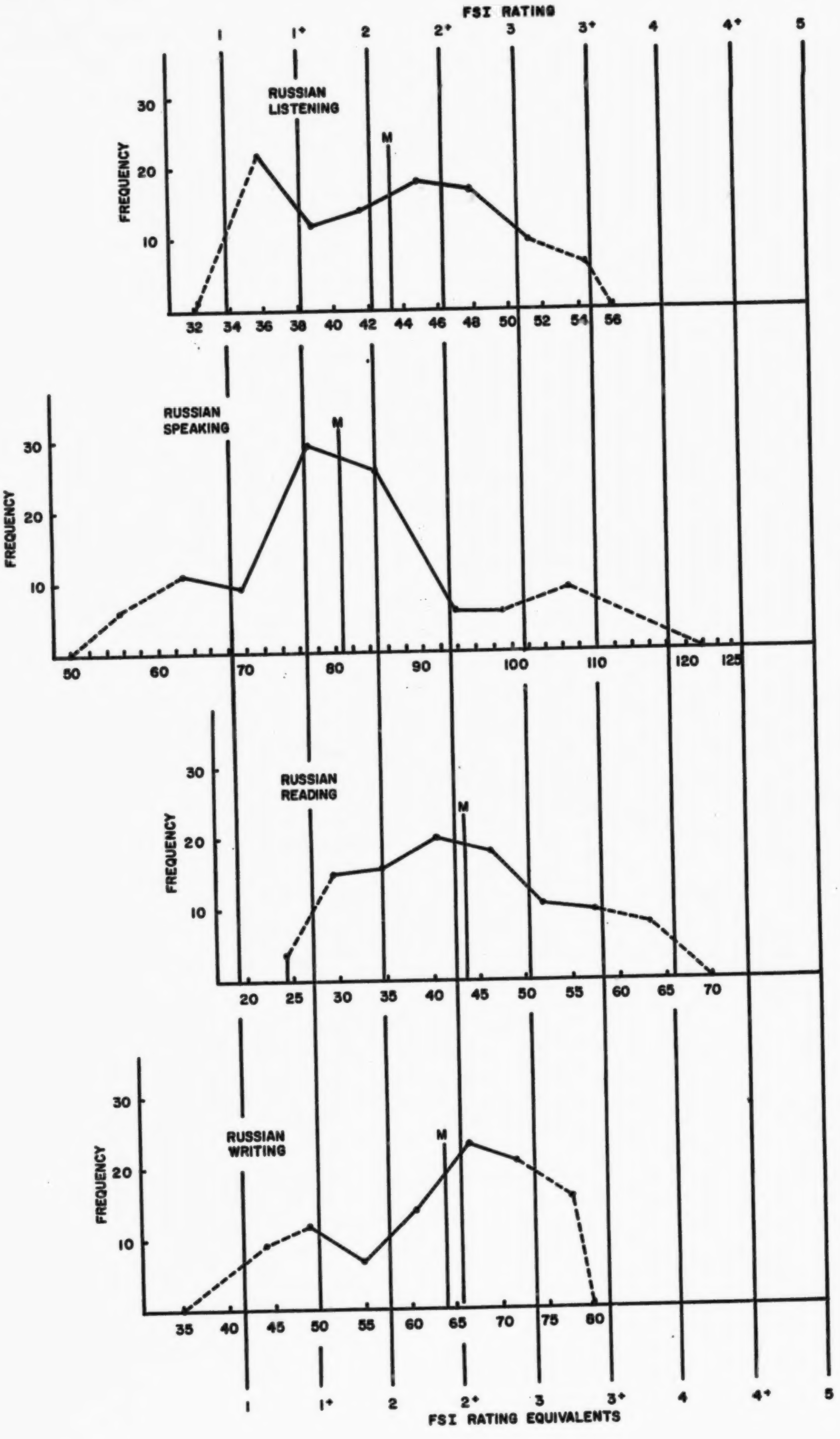
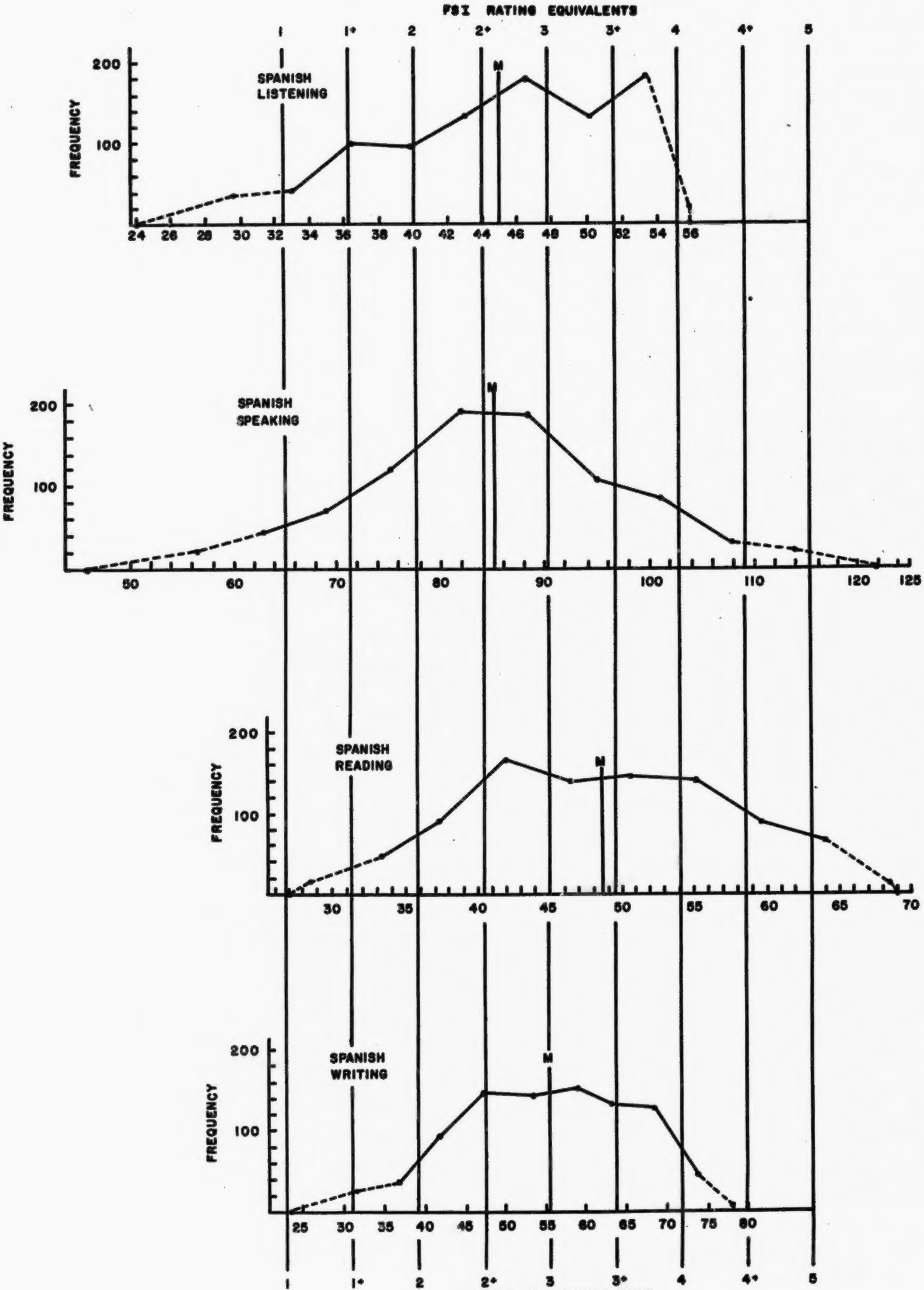


Figure 4.4

Frequency Distribution of Scores on MLA Skills Tests, Spanish, Tested Sample



Generally, for a negatively skewed distribution, the median will be greater than the mean.

For example, in appraising the general level of attainment in Listening achieved by the French group, the mean score is only 44.3, corresponding to an FSI rating a little below S-2+. The median (as estimated by a statistical formula that employs the mean and the measure of skewness g_1) is 45.4, more than a score point higher than the mean and appreciably closer to the presumed FSI equivalent of S-2+. Furthermore, it can be seen from the distribution (Figure 4.1) that there are substantial proportions of cases exceeding the FSI S-3 level, and if the test had been more difficult it is possible that many of those students shown as lying between S-3 and S-3+ would have scored at higher FSI rating equivalents.

Taking the results at their face value, however, we find that a general characteristic of the tested samples is that they are much poorer in Listening and Speaking skills than they are in Reading and Writing. This is true for all four languages except possibly for German, where at least the Listening median is in the range of the Reading and Writing medians according to the FSI equivalents. The Speaking test results are particularly low for the French group; in interpreting this finding, it must be borne in mind however, that the rating standards of the persons who judged the French speaking test results may have been for some reason unusually severe. (The footnote on page 12 suggested that rating standards of judges might vary from time to time.)

The distributions of Reading and Writing scores tend to be more normal (symmetric) in shape and to center around FSI ratings of R-3. Even for reading and writing skills, however, the results may be regarded as somewhat disappointing in comparison to what might be expected of foreign language majors at graduation from college. If the attainment of reading and writing skill at about the FSI R-3 level ("minimal professional skill") is taken to be one of the objectives of foreign language teaching at the college level, only about half of the seniors in our sample attained this objective, let alone such an objective in the listening and speaking skills.

2. Distributional statistics, "regular" and "odd" type cases

The means and standard deviations for "regular" and various "odd" type cases, in all languages except Italian,[1] are shown in Table 4.2. "Regular" cases, as explained previously, are those who began the study of their major foreign language at an introductory level at some point in the normal educational sequence--in elementary school, in secondary school, or in college--, and whose records showed no unusual language learning experiences, such as prolonged residence in a foreign country, that would classify them as "odd" types. Thus, the "regular" cases were intended to represent the typical native speaker of English who gets his foreign language training in the American educational system.

It may be seen by comparing data in Tables 4.1 and 4.2 that the "regular" cases showed mean skill scores slightly lower than those given for the total tested sample. This is mainly because many of the "odd" types excluded from the group had means that were considerably elevated because of their unusual language learning experiences, as will be pointed out below.

"Odd" type 4 cases, who listed no introductory courses in their backgrounds and who may therefore be presumed to have had unusual language learning experiences (e.g. learning from parents, special tutoring, etc.), did distinctly better on the tests than regular cases, on the average. There were substantial numbers of such cases in French, German and Spanish.

---

[1] For information on the test performance of students of Italian, see Table 4.3

Table 4.2

Distribution Statistics for MLA Skills Test Scores for "Regular" and "Odd" Type Cases[1]

		Listening		Speaking		Reading		Writing	
	N	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ
				FRENCH					
Regular cases	1115	44.12	7.63	73.56	9.25	48.81	8.59	47.16	9.54
Odd type 4	24	49.08	5.71	82.91	12.25	54.80	7.41	52.40	9.06
" " 6	5	51.60	1.87	73.80	8.96	52.40	9.57	53.20	11.87
" " 7	13-16	44.44	9.87	78.06	13.31	50.46	10.75	49.00	12.63
" " 8	9-10	48.30	7.52	82.11	10.29	52.11	12.41	51.90	12.47
" " 9	7	49.43	6.71	85.57	14.57	56.14	9.11	59.43	11.55
				GERMAN					
Regular cases	324	43.44	8.62	86.75	12.80	51.03	9.75	52.09	13.33
Odd type 4	27-29	51.90	5.34	104.96	11.83	59.68	8.39	65.11	9.16
" " 5	5	51.40	3.59	97.60	14.29	56.00	9.75	63.00	6.76
" " 6	4-5	49.80	5.33	102.50	12.53	60.00	7.07	62.40	4.08
" " 7	12-15	45.53	8.74	95.23	15.14	58.67	6.22	63.25	5.83
" " 8	2	54.00	--	102.50	--	63.00	--	61.00	--
" " 9	15-17	53.44	2.97	113.73	8.14	65.77	4.68	70.06	6.96
				RUSSIAN					
Regular cases	89	42.53	5.97	78.29	12.61	41.25	9.59	62.07	10.84
Odd type 3	3	44.00	--	93.33	--	56.67	--	68.33	--
" " 5	5-6	44.50	--	87.00	--	50.00	--	69.33	--
" " 7	1-2	46.00	--	120	--	49.50	--	69.00	--
" " 8	1	56.00	--	115	--	70	--	80	--
" " 9	3	51.33	--	109.33	--	59.33	--	70.33	--
				SPANISH					
Regular cases	816	44.47	6.79	83.99	12.01	47.83	8.41	54.62	10.13
Odd type 4	36	50.08	5.80	96.97	13.08	56.56	7.92	62.92	11.29
" " 5	1	43	--	82	--	45	--	45	--
" " 6	3-5	50.75	--	104.00	--	57.60	--	65.40	--
" " 7	18-20	45.58	6.70	85.89	14.20	47.90	6.67	54.89	9.36
" " 8	6	48.50	--	87.83	--	48.50	--	56.67	--
" " 9	24-26	52.69	2.19	105.29	11.07	63.69	3.74	71.81	3.56

---

[1] "Regular" cases are those whose records showed that they began the study of their major foreign language at an introductory level in either "grade school," secondary school, or college.

"Odd" type cases were classified as follows:

Type 4: First course listed was an intermediate or advanced college-level course, from which it may be presumed that their introduction to the language was not ordinary.

Type 5: Took training in the language during military service.

Type 6: Had spent two or more years in a country where the foreign language is the normal means of communication.

Type 7: Data on courses were missing, or so sparse that the case could not be otherwise classified.

Type 8: A miscellaneous category.

Type 9: Evidence from the questionnaire (often in the form of volunteered statements) indicated that the student was a native speaker of the language in which he was majoring.

Type 5 cases were those who indicated they had studied their major language in intensive courses given while they were in military service. Since there was no provision on the questionnaire for stating this, these comments were volunteered. There were 5 such cases in German, 6 in Russian, and 1 in Spanish. They did distinctly better on the skills tests, on the average, than the regular cases. This finding speaks well for the value of intensive language courses in military service. Of course, it could have happened that these students also had an opportunity to improve their language skills, after training, with practical work such as interpreting and translating.

Type 6 cases were those who volunteered the indication that they had spent two or more years residing in a country where the foreign language is the normal means of communication. There were 5 such cases in French, 5 in German, and 5 in Spanish. In general, they performed quite well on the tests, on the average, as compared to the regular cases. In German and Spanish, at least, their Speaking test scores were strikingly better than those of the regular cases. This finding agrees with a result to be discussed later in this report (Chapter VII), namely, that time spent abroad is an important variable in the attainment of foreign language skill.

Type 7 cases were those who could not be otherwise classified; for reasons unknown, they tended to perform somewhat better on the tests than regular cases.

Type 8 was a miscellaneous category reserved for those indicating various types of unusual language learning experiences not otherwise classifiable.

Particular interest attaches to the results for "odd" type 9, persons for whom there was evidence that they were native speakers of the language in which they were majoring. Through oversight, there was no provision in the questionnaire for indicating this, and therefore some cases of native speakers may have been included in the "regular case" category. Students were classified as type 9 only when there was clear evidence from the questionnaire, usually in the form of a volunteered statement, that they were native speakers. Of course, the definition of a native speaker presents some difficulties: many people could claim to be native speakers of a language learned in childhood and since forgotten.

At any rate, the data show 7 native speakers of French, 17 of German, 3 of Russian, and 26 of Spanish. We would expect such individuals to perform near the top of the score distributions, well in the range of FSI ratings of 4, 4+, and 5. (It should be remarked that the FSI rating of 5 is reserved for educated native speakers. We have no information on how the FSI procedure would rate native speakers with less than a university education.) The means for native speakers in Table 4.2 are in every case distinctly higher than those for regular cases; in fact, they are generally the highest among the means for the various "odd" type cases. The separation of the native speakers from the regular cases is particularly striking in the case of German and Spanish, where, in addition, the standard deviations tend to be much lower (indicating more homogeneity) than those of the regular cases. The French native speakers are, for some reason, not so distinctly superior.

It is only in German and Spanish that the native speakers have mean scores that are in the range of FSI ratings of 4, 4+, and 5. In French and Russian, the scores have FSI equivalents that tend to be more in the range of 3 and 3+. There are several possible reasons for the failure of the native speaker scores to have high FSI equivalents. First, the procedures employed for setting up the equivalents have many drawbacks: there were few if any native speakers involved in the equivalence experiment; there was in any case insufficient range of ratings and scores, and insufficient numbers of cases, to establish reliable equivalences; and the FSI rating procedure itself undoubtedly is not perfectly reliable or valid. Second, many of the tests appear to have ceilings; that is, they are not long enough, or difficult enough, to differentiate among the upper levels of language proficiency. Third, we do not have enough information about the "native speakers" in our sample to know whether they would be judged as native speakers by independent criteria. (Nevertheless, if a student claimed he was a native speaker of the language in which he was majoring, it is reasonable to assume that his knowledge of it was quite well formed through recent study even if he had once forgotten the language.) Finally, there is no guarantee that the MLA skills tests are sufficiently reliable and valid to identify native speakers with certainty.

3. Norms for Regular Cases

For certain purposes, it may be useful to have percentile norms for all the "regular" cases that were actually tested in the 1965 national study. These are presented in Table 4.3. They include, incidentally, information on the relatively small number of cases tested in Italian, and also data on the teacher preparation tests in the MLA battery (results of which are discussed in Chapter X). For norms estimated for the national sample after correction for sampling fractions, see Tables 5.11, 5.12, 10.3 , and 10.4 . The numbers of cases in Table 4.3 are slightly higher than those shown for regular cases in Table 4.2 because they include some cases whose data were otherwise incomplete.

4. Intertest correlations

Investigation of the correlations among the skills tests is of interest for the information it yields as to the extent to which these tests measure the same thing--overall language proficiency--or different things, i.e., language proficiency as exhibited in specific skills. The Pearsonian product-moment correlations among the four skills, for French, German, Russian, and Spanish, are shown in Table 4.4. Shown also, for comparison, are the intercorrelations reported (Educational Testing Service, 1961a, 1961b, 1962a, 1962b) for "post-test" scores of standardization samples of NDEA Institute teachers. The "post-test" scores were those obtained towards the end of the NDEA Institutes, i.e., after the teachers had received refresher training in the foreign language. These data were selected for comparison with the present data because they seemed more comparable, with respect to mean score levels, than the pre-test data (tests administered at the outset of the NDEA Institutes). For the present data, the numbers of cases are the same as those of Table 4.1; that is, they represent all cases ("regular" and "odd") that were complete on the four skills tests.

In view of the fact that the NDEA data appear to represent, judging by the standard deviations, a somewhat more heterogeneous sample than our sample, the correlations for the CFLT sample are not inconsistent with those obtained for NDEA Institute cases. That is, they are in general somewhat lower, but this is to be expected when the sample is more homogeneous.

All the correlations for the CFLT sample are highly significant in the sense that they depart significantly from what might be expected if the true correlations were zero. They are in fact almost uniformly high, high enough to indicate that the four skills tests measure, in the main, a single "common factor" of proficiency in the language concerned. Table 4.5 shows the proportions of test variance that can be accounted for by a single common factor, using a procedure described by Harman (1960, p. 122). Except in Russian, where the Listening test has the highest proportion of variance accounted for by a common factor, the Writing test is the test that has the highest proportion of common factor variance. It may be regarded, therefore, as the best single measure of overall language proficiency. The proportions of common variance in the tests range generally from about .70 to as high as .88, except for the Speaking test, where the common factor variances range from .53 to .72, due to a lower reliability of the Speaking test; partly it may be due also to a greater specificity of the skill measured by the Speaking test.

It should be noted, however, that these computations have not taken account of the possible effect of unreliability. Statistical procedures are available (Gulliksen, 1950, pp. 101-104) to correct correlations for the effects of unreliability, that is, to estimate what the correlations would be if the tests were perfectly reliable. This is often called the "correction for attenuation." However, it was not an objective of this investigation to determine the reliabilities of the tests. In order to perform the correction for attenuation, however, it is necessary to have estimates of the reliabilities of the tests applicable to the present sample. For this purpose we take the reliabilities

Table 4.3

Norms[1] for MLA Advanced FL Proficiency Tests Based on "Regular" Cases Tested, 1965 College Senior Major Samples

			FRENCH							GERMAN				
%ile Rank	List.	Speak.	Read.	Writ.	Appl. Ling.	Civ. & Cult.	%ile Rank	List.	Speak.	Read.	Writ.	Appl. Ling.	Civ. & Cult.	%ile Rank
99	56	107	70	75	67	67	99	56	115	69	77	65	70	99
97	55	93	66	69	63	59	97	56	110	68	74	64	67	97
95	55	90	64	64	61	58	95	55	107	67	72	62	66	95
90	54	86	61	61	58	55	90	54	103	64	69	59	64	90
85	53	83	59	59	56	53	85	53	102	61	67	58	61	85
80	52	81	57	56	55	52	80	52	98	60	65	56	59	80
75	51	80	56	55	54	50	75	51	96	59	63	55	57	75
70	50	78	54	53	52	49	70	50	94	57	61	54	55	70
65	49	77	53	52	51	48	65	49	92	56	60	53	54	65
60	47	76	52	50	50	47	60	48	91	55	58	52	53	60
55	46	75	51	49	49	47	55	46	89	53	56	51	51	55
50	45	74	49	47	48	46	50	45	87	52	54	50	50	50
45	44	73	48	48	47	45	45	44	86	50	51	48	49	45
40	43	72	47	45	47	44	40	43	84	49	48	47	48	40
35	41	70	46	43	46	43	35	40	83	48	46	46	47	35
30	40	69	44	42	45	43	30	38	81	45	44	45	46	30
25	38	67	43	40	44	42	25	36	79	44	42	44	44	25
20	37	66	41	39	42	41	20	34	76	42	40	44	43	20
15	35	64	39	37	41	40	15	33	73	40	37	42	42	15
10	33	62	37	34	39	38	10	31	69	38	34	41	40	10
5	30	58	34	31	37	36	5	29	66	34	30	38	37	5
3	29	56	33	29	36	34	3	28	62	32	26	37	37	3
1	27	52	31	26	33	32	1	27	57	30	24	36	35	1
No. Cases	1184	1123	1194	1193	445	460		338	327	337	338	116	111	
$\overline{X}$	44.000	73.568	48.774	47.044	48.661	46.104		43.618	86.817	51.095	52.183	49.879	50.928	
σ	7.666	9.265	8.614	9.573	7.235	6.457		8.583	12.808	9.680	13.224	7.278	8.752	

			ITALIAN (tentative)							RUSSIAN				
%ile Rank	List.	Speak.	Read.	Writ.	Appl. Ling.	Civ. & Cult.	%ile Rank	List.	Speak.	Read.	Writ.	Appl. Ling.	Civ. & Cult.	%ile Rank
99	55	102	55	68	58	58	99	56	112	65	80	69	78	99
97	51	99	54	66	57	56	97	53	107	60	78	64	71	97
95	48	97	53	65	56	55	95	52	101	59	77	61	64	95
90	46	94	52	64	55	54	90	51	96	55	75	57	63	90
85	45	93	49	63	54	53	85	49	89	52	73	56	62	85
80	44	92	48	60	54	52	80	48	87	49	72	55	60	80
75	43	91	47	59	53	51	75	47	85	48	70	52	57	75
70	43	91	46	58	53	50	70	46	83	46	69	51	56	70
65	42	90	45	57	52	49	65	45	81	44	68	50	55	65
60	42	90	44	56	52	49	60	44	80	43	67	50	54	60
55	42	89	43	55	51	48	55	43	79	42	66	49	53	55
50	41	87	42	54	49	47	50	43	78	40	65	49	52	50
45	41	85	41	53	47	46	45	42	77	39	63	47	52	45
40	40	84	40	52	47	44	40	41	76	37	61	46	51	40
35	40	83	40	51	46	44	35	39	75	36	59	46	50	35
30	38	81	39	50	46	43	30	38	74	35	56	45	49	30
25	37	78	38	49	46	43	25	37	72	34	54	45	48	25
20	33	76	37	48	45	42	20	36	69	32	52	44	46	20
15	32	72	37	46	44	40	15	35	65	30	51	44	45	15
10	31	70	34	42	44	38	10	34	60	29	49	43	43	10
5	29	65	29	37	43	35	5	33	56	27	40	42	41	5
3	27	62	28	33	43	32	3	33	53	26	38	40	39	3
1	24	57	27	25	42	28	1	31	50	24	35	37	36	1
No. Cases	24	24	24	23	8	8		92	91	92	90	18	20	
$\overline{X}$	39.750	84.833	42.292	53.304	49.375	46.125		42.304	78.198	41.033	61.956	49.111	52.800	
σ	5.613	10.893	6.931	8.728	4.947	6.451		6.087	12.846	9.676	10.950	6.253	8.171	

Table 4.3 (continued)

	SPANISH							PROFESSIONAL PREPARATION ALL LANGUAGES
%ile Rank	List.	Speak.	Read.	Writ.	Appl. Ling.	Civ. & Cult.	%ile Rank	
99	56	125	66	78	70	72	99	80
97	55	110	64	73	63	69	97	75
95	54	105	62	71	61	68	95	74
90	53	100	59	68	58	66	90	72
85	52	97	57	66	55	64	85	70
80	51	94	56	64	53	63	80	69
75	50	92	54	63	51	62	75	68
70	49	90	52	61	50	60	70	67
65	48	89	51	60	49	59	65	66
60	47	88	50	58	48	58	60	65
55	46	86	49	57	47	57	55	64
50	45	85	47	55	46	56	50	63
45	44	83	46	54	45	55	45	62
40	43	82	45	53	44	54	40	61
35	42	80	44	51	43	52	35	60
30	41	78	43	49	43	51	30	59
25	40	76	41	47	42	50	25	57
20	38	74	40	45	40	49	20	56
15	37	72	39	43	39	47	15	55
10	35	69	37	41	37	45	10	53
5	32	63	34	37	35	42	5	49
3	31	60	33	35	34	41	3	47
1	29	54	31	31	33	39	1	41
No. Cases	860	821	862	866	394	405		991
$\overline{X}$	44.442	83.962	47.773	54.510	46.680	55.536		62.126
σ	6.790	12.005	8.423	10.203	7.426	7.663		7.435

---

[1] For estimated norms for the national population, see Chapter V.

Table 4.4

Intercorrelations of MLA Skills Tests for CFLT National Sample, as Compared to Results for NDEA Institute Standardization Samples (ETS, 1961a, 1961b, 1962a, 1962b)

FRENCH

CFLT N = 1170 NDEA N = 1336*

	Listening	Speaking	Reading	Writing
Listening	1.00 [.844]	.68 (.800)	.73 (.800)	.75 (.797)
Speaking	.68 (.800)	1.00 [.825]	.58 (.736)	.65 (.782)
Reading	.73 (.800)	.58 (.736)	1.00 [.883]	.80 (.858)
Writing	.75 (.797)	.65 (.782)	.80 (.858)	1.00 [.924]
Mean (Form A)	44.3 (42.0)	73.9 (83.6)	49.0 (47.4)	47.4 (46.9)
S.D.	7.6 (8.4)	9.5 (17.1)	8.6 (10.5)	9.7 (12.5)

GERMAN

CFLT N = 388 NDEA N = 297*

	Listening	Speaking	Reading	Writing
Listening	1.00 [.844]	.78 (.725)	.85 (.840)	.85 (.806)
Speaking	.78 (.725)	1.00 [.806]	.78 (.700)	.80 (.771)
Reading	.85 (.840)	.78 (.700)	1.00 [.900]	.87 (.864)
Writing	.85 (.806)	.80 (.771)	.87 (.864)	1.00 [.924]
Mean (Form A)	44.8 (43.0)	89.8 (85.5)	52.7 (48.7)	54.3 (51.3)
S.D.	8.6 (7.0)	14.4 (12.9)	10.1 (11.1)	13.5 (16.4)

RUSSIAN

CFLT N = 102 NDEA N = 176*

	Listening	Speaking	Reading	Writing
Listening	1.00 [.798]	.69 (.746)	.81 (.766)	.78 (.754)
Speaking	.69 (.746)	1.00 [.786]	.68 (.708)	.68 (.666)
Reading	.81 (.766)	.68 (.708)	1.00 [.850]	.75 (.705)
Writing	.78 (.754)	.68 (.666)	.75 (.705)	1.00 [.914]
Mean (Form A)	43.3 (39.8)	80.8 (74.8)	43.3 (39.4)	63.4 (57.9)
S.D.	6.2 (5.8)	14.8 (17.0)	11.2 (10.6)	11.2 (14.5)

SPANISH

CFLT N = 900 NDEA N = 1334*

	Listening	Speaking	Reading	Writing
Listening	1.00 [.860]	.67 (.770)	.77 (.800)	.77 (.803)
Speaking	.67 (.770)	1.00 [.789]	.61 (.702)	.65 (.724)
Reading	.77 (.800)	.61 (.702)	1.00 [.877]	.81 (.862)
Writing	.77 (.803)	.65 (.724)	.81 (.862)	1.00 [.902]
Mean (Form A)	45.0 (42.4)	85.2 (79.8)	48.6 (44.5)	55.5 (49.2)
S.D.	6.9 (7.8)	12.8 (16.6)	8.8 (9.7)	10.5 (12.6)

*Forms A and B were spiraled between or within institutes and between pre-test and post-test. The correlations in square brackets [ ] are between pre-test and post-test scores, regardless of form. The correlations in parentheses ( ) are between-skills correlations for post-test scores, regardless of form. The means and standard deviations for NDEA data are for Form A post-test scores, with N's approximately half those for the correlations. NDEA post-test data were selected for comparison because they seemed more comparable than pre-test data to the CFLT data.

Table 4.5

Proportions of Common Factor Variance
Accounted for by the MLA Skills Tests*
(Derived from Correlations for the CFLT Sample in Table 4.4)

Test	FRENCH	GERMAN	RUSSIAN	SPANISH
Listening	.7655	.8361	.8194	.7849
Speaking	.5322	.7223	.5986	.5280
Reading	.7074	.8564	.7760	.7596
Writing	.8078	.8778	.7456	.8052

*See Table 4.6 for estimates after correction for attenuation.

of the tests given for the NDEA standardization sample (Educational Testing Service, 1964) and estimate reliabilities for the present sample by taking account of the differences in standard deviations. The results, using a formula given by Gulliksen (1950, p.111, formula 5), are shown in Table 4.6, which also shows correlations corrected for attenuation and proportions of variance attributable to a common factor after correction for attenuation. Data from the NDEA standardization sample do not show reliabilities for any of the Speaking tests, but the leaflet published by Educational Testing Service indicates that in certain experiments (with what languages we are not told) an interscorer reliability of .89 was obtained for the Speaking test. This figure, therefore, was used for all Speaking tests in the present computations.

Several points are of note concerning Table 4.6. First, the generally somewhat lower reliability coefficients for the CFLT sample, as compared to those given for the NDEA Institute standardization samples, are evidently due to the generally smaller variance of the CFLT sample. The reliability of the French Speaking test is estimated as only .64 for the CFLT sample. (Of course, the standard error of measurement is assumed to be constant over samples.) This explains the somewhat depressed correlations for the Speaking test in Table 4.4. After the correlations are corrected for attenuation, as shown in Table 4.6, it would appear that the various skills tests are even more strongly associated; that is, they are shown even more to measure a single common factor of language proficiency. The estimated proportions of common factor variance shown in Table 4.6 range from .7042 to .9338. For all languages except Russian, the Writing test maintains its superiority as the best measure of that common factor, and except in French, the Speaking test is least associated with the common factor.

It is not surprising that the four skills tests should be found to measure primarily a single factor of language proficiency in common. Basic competence in a language--knowledge of its phonology, morphology, syntax, and lexicon--is required by each of the tests, no matter what particular "skill" it measures. The high loading of the Writing test on the common factor may reflect the fact that this test is probably most demanding with respect to the morphology and syntax of the language. Many of the other tests appear to demand knowledge primarily of lexicon, which some would regard as less close to the heart of language structure.

The fact that the Speaking test is least associated with the common factor of overall language proficiency may indicate that the requirements of the task set by this test are fairly specific and possibly to some extent unrelated to the measurement of language proficiency. The examinee may be unaccustomed to speaking into a microphone and his perform-

Table 4.6

Estimated Reliability Coefficients for MLA Tests Applied to CFLT Sample and Estimated Correlations Corrected for Attenuation (Derived from Data of Table 4.4)

	NDEA Reliability	Est.'d Reliability, CFLT	Correlations Corrected for Attenuation				Est.'d Proportions of Common Factor Variance After Attenuation Correction
			List.	Speak.	Read.	Writ.	
				FRENCH			
Listening	.91	.89	(1.00)	.90	.81	.83	.8601
Speaking	.89*	.64	.90	(1.00)	.76	.85	.8316
Reading	.93	.90	.81	.76	(1.00)	.88	.7860
Writing	.94	.90	.83	.85	.88	(1.00)	.8975
				GERMAN			
Listening	.90	.93	(1.00)	.85	.92	.91	.8990
Speaking	.89*	.91	.85	(1.00)	.85	.87	.7937
Reading	.94	.93	.92	.85	(1.00)	.93	.9209
Writing	.97	.94	.91	.87	.93	(1.00)	.9338
				RUSSIAN			
Listening	.86	.88	(1.00)	.80	.90	.87	.9311
Speaking	.89*	.85	.80	(1.00)	.76	.76	.7042
Reading	.92	.93	.90	.76	(1.00)	.81	.8344
Writing	.96	.93	.87	.76	.81	(1.00)	.8017
				SPANISH			
Listening	.91	.89	(1.00)	.78	.87	.86	.8819
Speaking	.89*	.82	.78	(1.00)	.71	.75	.6439
Reading	.91	.89	.87	.71	(1.00)	.90	.8535
Writing	.94	.91	.86	.75	.90	(1.00)	.8848

*Assumed reliability for all tests on the basis of a report of interscorer reliability found in an experiment done at ETS (ETS, 1964 a, p. 30).

ance may depend not so much on his language proficiency as on his general fluency and self-confidence in spontaneous speech--whether in his native language or a foreign language.

Linearity of intertest regressions

Each Pearsonian product-moment correlation coefficient shown in Table 4.4 is actually the slope of the straight line of best fit to the regression of one variable on another variable (both variables being in standardized form). Interpretation of such correlation coefficients must be qualified by the assumption that a straight line is (aside from sampling fluctuations) the best fit to the data. Often, a curved line is a significantly better fit to the regression, in which case prediction of one variable from another should take account of this. Furthermore, curvilinear regressions are often an indication that a given test does not have a uniformly spaced scale of measurement; for example, that it has a "ceiling" or a "floor." We have already adduced evidence that some of the MLA skills tests have ceilings, in that they fail to measure adequately the upper levels of achievement. A finding of a curvilinear relationship between test variables could also indicate some specially interesting functional relation between the underlying variables.

Out of the 48 possible combinations of MLA skills tests language by language, there were only 6 instances of significant departures of regressions from linearity, four significant at the 1% level and 2 others significant at the 5% level; these instances are detailed in Table 4.7. Although one or two of these instances may be significant only by chance, they are all worth examining. The best way of doing this is to plot the regression lines graphically, and consequently we present Figures 4.5 through 4.9.

Some explanation is due. Each figure presents a plot of the means for the ordinate variable (the variable on the vertical scale) for given values of the abscissal variable (the variable on the horizontal scale). For example, in Figure 4.5, the French Listening score means are plotted against given values for the French Speaking test score. The point at the left lower extreme indicates that the average value of the French Listening score is 33.4 for all cases with scores on the French Speaking test grouped near 50.7; the next point indicates a mean Listening score of 34.0 for cases grouped near 57.5; and so on. The points do not form a straight line, as they would, approximately, if the regression were linear.

Figure 4.5 suggests that measurement on the French listening test is attenuated at both extremes of its distribution; that is, on the assumption that there is valid measurement at the extremes of the Speaking scale, such measurement is not well reflected in the Listening test scores. The regression line tends to flatten out both at the bottom and the top of the curve. At the bottom, it approaches a chance score, and at the top, it approaches the maximum score. Thus, we may conclude that the French listening test has an insufficient range of difficulty for the target population: it does not provide enough "floor" for low ability students, or enough "ceiling" for the students of advanced levels of achievement. This conclusion is supported also by the results depicted in Figure 4.6, where it is seen that there is some bending of the regression line as it approaches the maximum possible Listening score. There appears to be no significant bending at the bottom, to be sure, but it is possible that the lower end of the Writing test is also insufficiently discriminative.

A similar phenomenon is to be observed for the Listening test results in German, Russian, and Spanish in Figures 4.7, 4.8, and 4.9, respectively. In Figure 4.7, for German, we again may make the assumption that the Speaking test score is adequately discriminative at the extremes of its distribution, and on this basis we conclude that the German Listening test has a low ceiling whereby students at advanced levels are not adequately discriminated. There is also some suggestion that the German Listening test does not measure adequately at the lower end of the scale, either. In Figure 4.8, for Russian, the ordinates are reversed: for some reason, the significant departure from linearity is observed for the regression of Speaking on Listening rather than for the regression of Listening on Speaking, as was the case in French and German. Nevertheless, the net effect is the same: large differences at the upper end of the Speaking test score fail to be

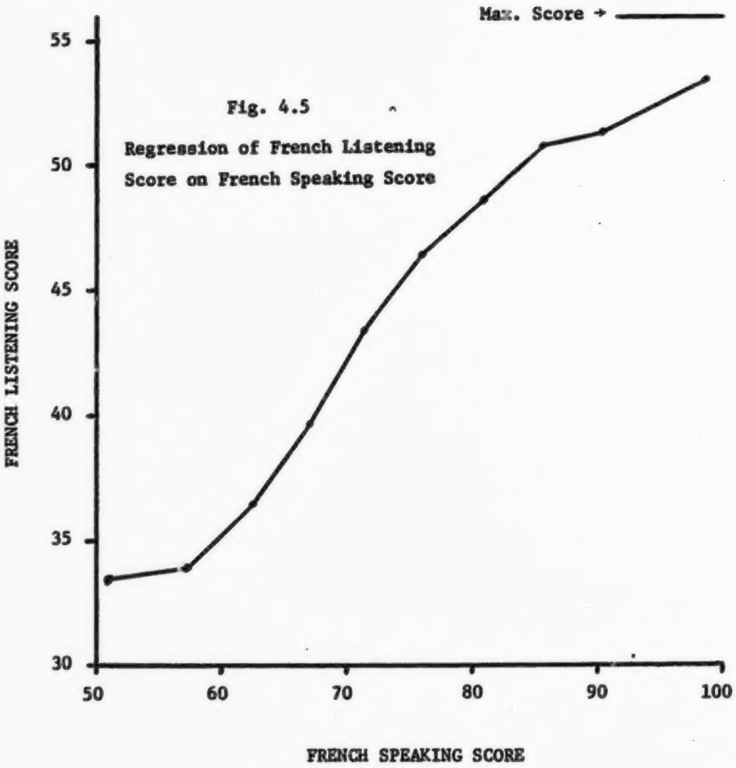


Fig. 4.5

Regression of French Listening Score on French Speaking Score

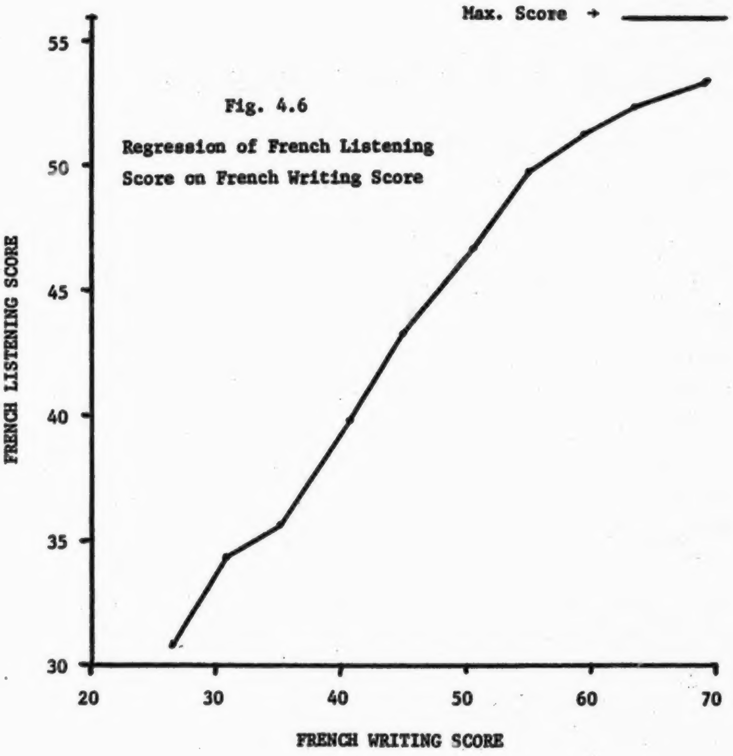


Fig. 4.6

Regression of French Listening Score on French Writing Score

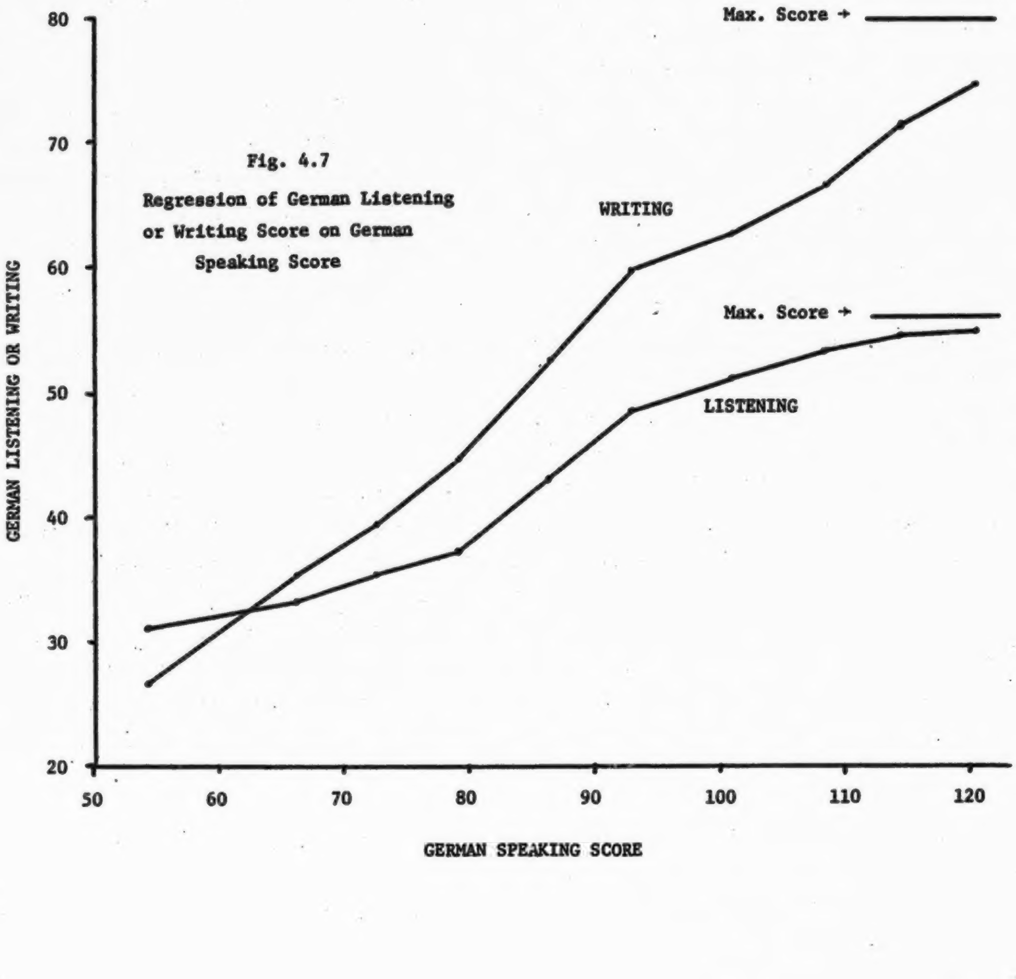


Fig. 4.7

Regression of German Listening or Writing Score on German Speaking Score

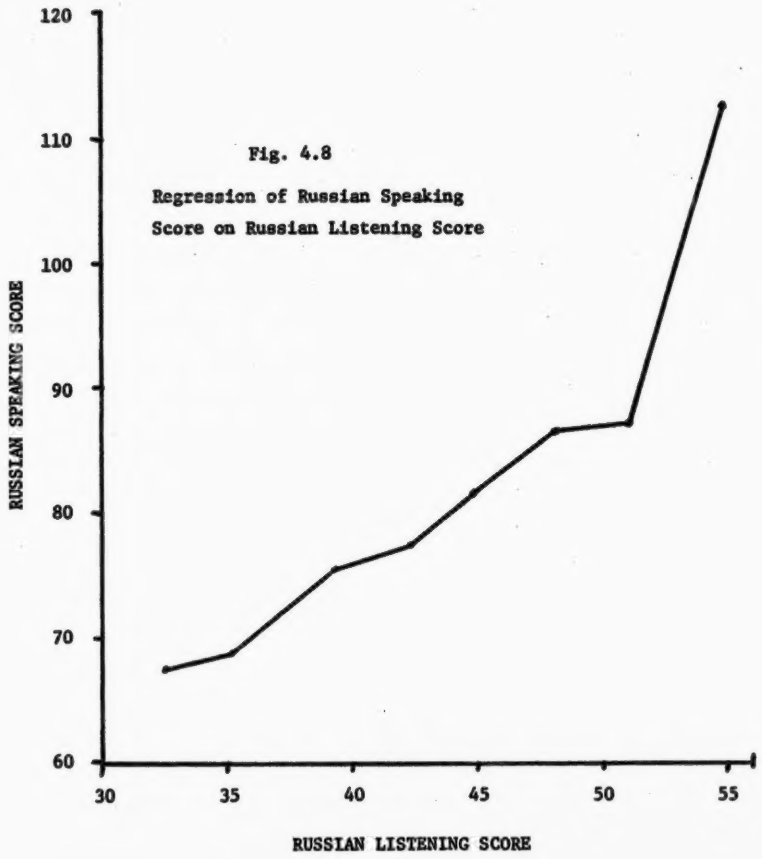


Fig. 4.8

Regression of Russian Speaking Score on Russian Listening Score

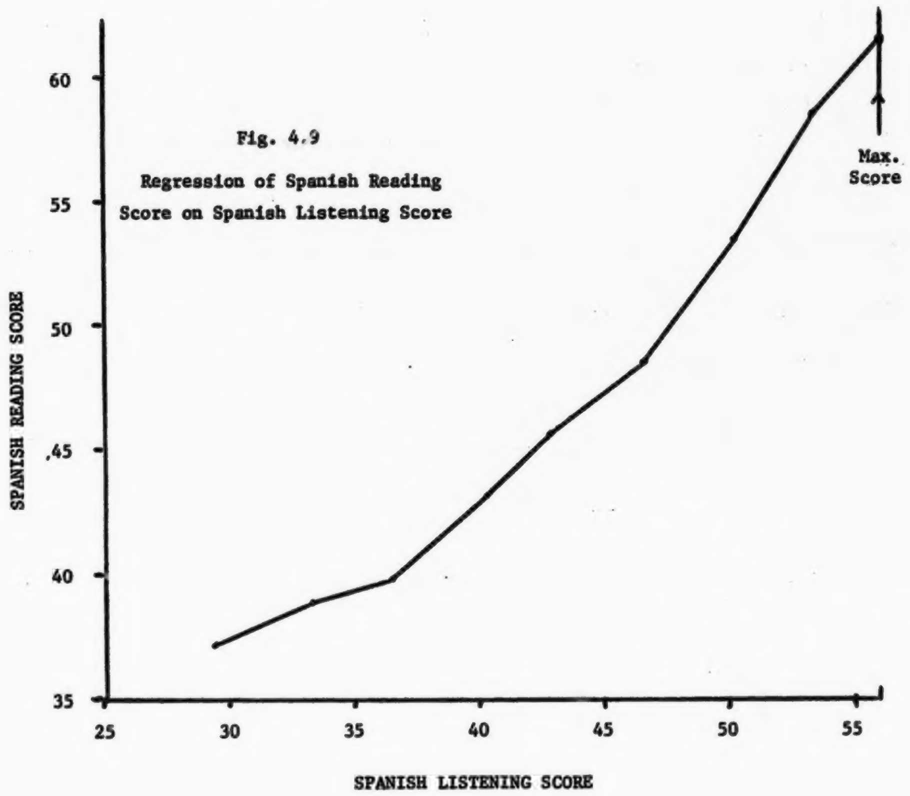


Fig. 4.9

Regression of Spanish Reading Score on Spanish Listening Score reflected in correspondingly-sized differences on the Listening test, suggesting that the Russian Listening test is not sufficiently difficult. In Spanish, the low ceiling of the Listening test exhibits itself when the regression of the Reading test on Listening is examined. At the same time the flattening of the regression line at the lower end suggests that the Spanish Reading test does not have enough "floor" to accommodate low ability students who are, nevertheless, differentiated by the Listening test.

Figure 4.7 suggests also that the German Writing test may not have enough ceiling to differentiate advanced students who are, nevertheless, differentiated by the Speaking test.

The results here are entirely consistent with the observations made in Section 1 of this chapter that the Listening tests in all languages fail to cover enough territory on the FSI scale. It will be recalled from that Section that Speaking tests were generally adequate in their coverage of the FSI scale, however, and this fact tends to justify the assumption that has been made here about the Speaking tests, namely, that they do provide adequate differentiation among students at either extreme of the scale.

The observations made in this chapter concerning test characteristics could well be useful in preparing revised or further forms of the MLA skills tests.

Table 4.7

Instances of Significant Departure from Linearity

in Intercorrelations of MLA Skills Tests

Language	N	Independent Variable	Dependent Variable	r	(Eta) η	F-ratio
FRENCH	1170	Speaking	Listening	.68	.69	4.9**
		Writing	Listening	.75	.76	3.2**
GERMAN	388	Speaking	Listening	.78	.80	4.6**
		Speaking	Writing	.80	.81	2.5*
RUSSIAN	102	Listening	Speaking	.69	.74	2.7*
SPANISH	900	Listening	Reading	.77	.79	4.8**

$^{**}p < .01$; $^{*}p < .05$

Chapter V

ESTIMATES OF FOREIGN LANGUAGE SKILL ATTAINMENTS IN THE TOTAL POPULATION

The previous chapter reported results actually obtained from the stratified sample of college senior foreign language majors obtained through a nationwide testing program. It was pointed out, however, that the sample may have contained some bias due to non-response of institutions and students. This chapter attempts to estimate the amount of this bias, at least that due to student non-response. Furthermore, it presents national norms estimated by a process of correcting for variation in the representation of the different strata in the sample.

1. Bias due to student non-response

We take up the matter of student non-response first because any corrections for it would have to be made prior to corrections for variations in the representation of strata. Student non-response is defined as the failure of a graduating FL major at a selected institution to volunteer to take any of the tests.

It was pointed out in Chapter III, Section 2 that student response rates were far short of the desired 100% that had been planned for in the sample design. Student non-response was particularly poor in the larger institutions, but it was nowhere near perfect in the smaller institutions. There were few institutions where 100% of the available students showed up for the testing.

Student non-response would not be a matter of any great concern in interpreting national norms if it could be shown that it was not correlated with test performance. That is, if tested and non-tested students do not differ in average test performance, the statistics obtained for the tested students would be good estimates of the statistics for all students--i.e., the tested pooled with the non-tested. It was decided to make a special study to determine whether the students actually tested could be considered to be a representative sample of all graduating language majors. On intuitive grounds, there was the possibility that the "better" students were more likely to volunteer for testing than "poorer" students.

But how were the "better" students to be identified independently of the test results? The only possibility that suggested itself as feasible was the use of foreign language grades, available on college transcripts, as independent measures of foreign language competence. Although such a measure was probably the best available under the circumstances, it cannot by any means be regarded as an adequately valid and reliable measure. College grades are notoriously unreliable and subject to shifting standards. Furthermore, they do not necessarily reflect to any great extent the foreign language competence levels presumably measured by the skills tests in the MLA Proficiency Test battery. Grades in foreign language literature courses, for example, may have little to do with any aspect of the student's competence in the foreign language; they may reflect his sensitivity to literature, his general ability to write critical essays (whether in English or the foreign language), the amount of effort he puts in study, or his ability to impress the instructor. For many of the students in our project, the only grades in foreign language available would be those in literature and other courses not directly addressed to the promotion of foreign language skills.

It was decided, nevertheless, to obtain the transcripts of all class of 1965 foreign language majors at selected institutions in order to see whether tested students tended to be those with higher foreign language grades than students who did not volunteer to take the tests, and further, whether foreign language grades tended to be correlated with MLA Proficiency Test skill scores. If grade-point-average differences were found between tested and non-tested students, and if there were appreciable correlations between grades and test scores in the case of the tested students, it would be possible, through certain statistical procedures, to estimate the amount of bias in the test score statistics due to student non-response.

For this purpose, 15 medium-large institutions (7 public, 8 private) that had reported from 21 to 74 graduating majors in all languages and that tested anywhere from 25 to 90 per cent of students were selected (hereafter, this is called the "bias study sample"). They were representative of all parts of the country and of many types of institutions. By correspondence and/or telephone, registrars of these institutions were asked to send transcripts of all students who had been enrolled as foreign language majors in the class of 1965. Confidentiality of the individual transcripts was promised, and all registrars agreed to provide them. (Most registrars charged a fee for this service.)

On receipt of the transcripts, the project staff computed a grade-point-average in the relevant foreign language courses for each student. (That is, only the courses in the student's major language were counted; in the case of the students tested in the CFLT project, this was also the language in which they took the MLA tests.) Letter grades were converted to numerical codes as follows: A=5, B=4, C=3, D=2, E or F=1. Pluses and minuses were ignored since many schools used only letters with no pluses and minuses attached.

A preliminary statistical analysis was performed to see whether there were significant differences between the foreign language GPA's of tested and non-tested students. Results are shown in Table 5.1. Over all schools, there was a highly significant difference favoring the tested students. This result is probably reliable even though grading standards may have varied from institution to institution, for the trend favors the tested students in all but two schools. The latter two schools were those with the highest percentages of tested students; there may be some meaning in the fact that the non-participating students at these institutions were on the average the best students. In 6 of the 13 institutions where the GPA's favored the tested students, the differences were significant at the 5 per cent level or beyond. One may conclude that there was a definite trend whereby the students who volunteered to take the tests were generally the better students, and this trend probably existed over the whole sample of institutions participating in the nationwide study.

It may be noted, incidentally, that the average GPA's for tested and non-tested students correspond to letter grades in the A to B range in nearly all cases.

Whether the trend noted above actually constituted a bias in the results of the MLA tests was less clear, however. To investigate this question, correlations between MLA test scores and foreign language grades were studied for the tested students alone. First, however, all GPA's within institutions were converted to standard scores such that all students within an institution would have, on the average, the same mean and standard deviation. (For convenience, standard scores with a mean of 10.00 and a standard deviation of 1.00 were derived.) This was done with a view to eliminating, as much as possible, any institutional variability in grading standards that may have been present. Next, all tested cases in each language were assembled for separate analysis because it could not be assumed that the MLA test scales were comparable across languages. In all, there were 124 tested cases in French (49.4% of 251 available), 40 in German (48.2% of 83), 8 in Russian (34.8% of 23), and 101 in Spanish (50.5% of 200). There were no cases in Italian in these 15 institutions. In each language group, correlations were determined between MLA test scores and the standardized GPA's, with results as shown in Table 5.2. Because of incomplete data on some of the MLA tests, however, these correlations were based on only 105 cases in French, 36 in German, 8 in Russian, and 84 in Spanish. Except in Russian, where the range of GPA's was very restricted and the average GPA was exceptionally high, the correlations between MLA skill scores and GPA's were all posi-

Table 5.1

Comparison of Foreign Language GPA's of Tested and Non-Tested Students at Fifteen Selected Institutions, in Order of Percentage Tested

Institution Code	N's			% Tested	Tested	Not Tested	All	t	p
	Tested	Not Tested	Total						
035	9	27	36	25	4.174	4.024	4.062	0.8793	n.s.
610	10	27	37	27	4.765	4.222	4.369	2.1818	< .05
665	13	31	44	30	4.332	3.949	4.062	2.5275	< .05
685	23	51	74	31	4.345	4.069	4.155	2.0836	< .05
860	9	16	25	36	4.711	4.231	4.438	2.1583	< .05
205	15	25	40	38	4.305	4.167	4.219	0.9819	n.s.
620	14	22	36	39	4.346	4.119	4.207	1.9749	n.s.
385	11	15	26	42	4.366	4.142	4.237	0.9844	n.s.
730	13	14	27	48	4.290	3.850	4.062	2.0575	n.s.
457	23	22	45	51	4.123	4.049	4.087	0.4593	n.s.
655	15	8	23	65	4.387	3.827	4.192	2.8351	< .01
287	29	13	42	69	4.126	3.964	4.076	1.1057	n.s.
755	17	4	21	81	4.254	3.602	4.130	2.6737	< .05
422	37	5	42	88	4.284	4.344	4.291	-0.2896	n.s.
247	35	4	39	90	4.192	4.587	4.233	-1.6067	n.s.
All	273	284	557	49	4.289	4.092	4.189	4.7369	<.001

tive, ranging from .27 to .55 with a median at .37. Since the range of actual GPA's was relatively small, these correlations probably suggest that there is indeed a substantial relationship between foreign language GPA and ability as measured by the MLA skills tests. The correlations were judged large enough (except in Russian) to justify further statistical computations to estimate the amount of bias in MLA test results introduced by the less-than-perfect student response rate.

In these computations, use was made of certain theoretical relationships between correlational data obtained on two samples with different dispersions of scores on two correlated variables. These relationships are often used in "correcting" correlations for "restriction of range of talent." In the present case, however, they were used not to correct the correlations but to estimate the means and standard deviations of MLA test scores for all cases if scores had been available for non-tested students as well as tested students. The basic assumption made in these computations is that when a sample is "explicitly" selected from a larger sample in such a way that the standard deviation of a certain variable X in the smaller sample (s_x) is different from the standard deviation in the larger sample (S_X), this will not change (except for sampling fluctuations, which are ignored) the parameters of the linear regression equation for predicting values of a correlated variable Y; that is, regardless of whether values for the larger or smaller sample are used,

$$\hat{Y} - \overline{Y} = R_{XY} \frac{S_Y}{S_X} (X - \overline{X}) = r_{xy} \frac{s_y}{s_x} (X - \overline{X}) ,$$

where Y is the predicted value of Y; $\overline{Y}$ and $\overline{X}$ are, respectively, the means of Y and X; R_{XY} and r_{xy} are respectively the correlations of X and Y in the larger and the smaller sample; and S_Y and s_y are, respectively, the standard deviations of Y in the larger and the smaller sample.

In the present case, the Grade-Point-Average (GPA) is the variable subject to "explicit" selection (that is, the better students volunteered to take the tests) and may be designated X, while the MLA skill test score is the variable subject to "incidental" selection through its correlation with X, and may be designated Y.

From the data available from the bias study sample, we know, in any given case, $\overline{X}$, $\overline{x}$ (the mean GPA in the tested group), S_X, s_x, r_{xy}, $\overline{y}$ (the mean MLA skill test score in the tested group), and s_y. The problem is to estimate $\overline{Y}$ and S_Y, the mean and standard deviation of Y if all students had been tested.

Now, $\overline{y}$ can be regarded as the value of Y that would be predicted by the regression line from $\overline{x}$ as a selected value of X, since it is the mean MLA skill score for the tested group. Rewriting the regression equation (using the values known for the tested group), we have:

$$\overline{y} - \overline{Y} = r_{xy} \frac{s_y}{s_x} (\overline{x} - \overline{X}) ,$$

and we can immediately solve for $\overline{Y}$ by the equation

$$\overline{Y} = \overline{y} - r_{xy} \frac{s_y}{s_x} (\overline{x} - \overline{X}) ,$$

in whose right-hand side all values are known.

To estimate S_Y, we can use formula (20) given by Gulliksen (1950, p. 138), whose notation is used here:

$$S_Y = s_y \sqrt{1 - r_{xy}^2 + r_{xy}^2 (S_X^2/s_x^2)} \ .$$

It is also possible to estimate R_{XY}, the correlation between MLA skill score and GPA in the larger sample, by Gulliksen's formula (18):

$$R_{XY} = \frac{S_X r_{xy}}{\sqrt{S_X^2 r_{xy}^2 + s_x^2 - s_x^2 r_{xy}^2}}$$

However, this is of little interest in the present case.

By use of the data in Table 5.2, the estimated means and standard deviations of MLA skill test scores for the complete comparability samples in French, German, and Spanish were computed by the above formulas; the results are shown in Table 5.3. The table indicates also: the amount of change in the means effected by the estimation procedure; the means and standard deviations of all "regular cases" actually tested in the national sample (data from Table 4.3); and means and standard deviations of the national sample adjusted for student non-response bias by shifting the means by the amount of change found in the comparability study and inflating (or deflating) the standard deviations by the ratio of the two S.D.'s in the comparability study. What is of chief interest here is the fact that the amounts of change (downward) effected by the estimation procedure are almost all less than one converted score point, even for variables with relatively large standard deviations. From this we can conclude that the amount of bias in our data due to student non-response is relatively small. For practical purposes, any national norms derived from the data on tested students can be corrected for student non-response bias very easily by simply shifting converted scores by one point, downward.

One caution here, of course, is that the correction procedure used here depends upon the correlations between MLA skill test scores and foreign language grade-point-averages. If grade-point-averages had correlated higher than they did with MLA skill test scores, the corrections for student non-response would have been greater in magnitude. If d is the amount of shift actually computed, the maximum possible shift that could have occurred if the MLA-GPA correlations had been perfect would be d/r_{xy}.

It will be noticed in Table 5.3 that the bias study sample, even when examined in terms of estimated MLA skills test means for the total sample, was in nearly every case superior to the national CFLT sample. This probably reflects the fact that the bias study sample was assembled from data from the larger institutions; it will be shown later (Chapter V, Section 2) that students at the larger institutions tend to have somewhat higher MLA skills test scores than those at the smaller institutions.

Table 5.2

Correlations among MLA Skills Test Scores and between MLA Skills Test Scores and Standardized FL Grade-Point-Average, by Language, for Cases with Complete Test Scores

(MLA-GPA Bias Study Sample at 15 Selected Institutions)

		MLA Skill Test Scores				Standardized G P A's (Mean=10, S.D.=1)	Standardized GPA's All tested	All cases
		L	S	R	W			
					FRENCH	(N = 105)	(N=124)	(N=251)
Mean		45.99	76.11	51.88	49.88	10.13	10.20	9.94
S.D.		7.53	9.11	8.02	9.12	.97	.94	1.02
	L	1.00	.68	.70	.73	.41	-	-
r's	S	.68	1.00	.55	.71	.41	-	-
	R	.70	.55	1.00	.75	.51	-	-
	W	.73	.71	.75	1.00	.47	-	-
					GERMAN	(N = 36)	(N=40)	(N=83)
Mean		45.86	89.89	54.97	57.92	10.52	10.50	10.27
S.D.		8.89	13.46	10.10	13.36	.90	.90	.96
	L	1.00	.76	.88	.85	.27	-	-
r's	S	.76	1.00	.77	.79	.29	-	-
	R	.88	.77	1.00	.90	.34	-	-
	W	.85	.79	.90	1.00	.33	-	-
					RUSSIAN	(N=8)	(N=8)	(N=23)
Mean		44.63	81.50	43.00	68.50	10.99	10.99	10.68
S.D.		4.47	8.33	5.90	5.71	.69	.69	1.15
	L	1.00	.11	.41	.70	-.23	-	-
r's	S	.11	1.00	-.08	.23	-.25	-	-
	R	.41	-.08	1.00	.70	-.25	-	-
	W	.70	.23	.70	1.00	-.18	-	-
					SPANISH	(N=84)	(N=101)	(N=200)
Mean		45.57	89.01	48.02	55.81	10.03	10.04	9.88
S .D.		6.57	9.56	8.36	8.60	.95	.96	.94
	L	1.00	.53	.77	.62	.27	-	-
	S	.53	1.00	.40	.46	.31	-	-
	R	.77	.40	1.00	.76	.39	-	-
	W	.62	.46	.76	1.00	.55	-	-

Table 5.3

Study of Bias Due to Student Non-Response, and Estimates of Corrections for National Sample Data, for French, German, and Spanish Students

MLA Skill Test	Bias Study Sample					National CFLT Sample (Regular Cases)				
	Tested students		Estimated for Total Sample		Shift in	Actual Data			Corrected for Non-Response(Est.)	
	Mean	S.D.	Mean	S.D.	Mean	N	Mean	S.D.	Mean	S.D.
					FRENCH					
	(N = 105)		(N = 251)							
Listen.	45.99	7.53	45.39	7.59	-.60	1184	44.00	7.67	43.40	7.73
Speak.	76.11	9.11	75.39	9.19	-.72	1123	73.57	9.27	72.85	9.35
Reading	51.88	8.02	51.09	8.12	-.79	1194	48.77	8.61	47.98	8.71
Writing	49.88	9.12	49.05	9.22	-.83	1193	47.04	9.57	46.21	9.68
					GERMAN					
	(N = 36)		(N = 83)							
Listen.	45.86	8.89	45.20	8.93	-.66	338	43.62	8.58	42.96	8.61
Speak.	84.89	13.46	88.82	13.54	-1.07	327	86.82	12.81	85.75	12.89
Reading	54.97	10.10	54.03	10.18	-.94	337	51.10	9.68	50.16	9.76
Writing	57.92	13.36	56.71	13.45	-1.21	338	52.18	13.22	50.96	13.31
					SPANISH					
	(N = 84)		(N = 200)							
Listen.	45.57	6.57	45.29	6.57	-.28	860	44.44	6.79	44.16	6.79
Speak.	89.01	9.56	88.55	9.55	-.46	821	83.96	12.01	83.50	12.00
Reading	48.02	8.36	47.51	8.35	-.51	862	47.77	8.42	47.26	8.41
Writing	55.81	8.60	55.07	8.58	-.74	866	54.51	10.20	53.77	10.18

2. Variations in Foreign Language Skill Associated with Stratification Variables in the Sample Design

Implicit in the sample design was the assumption that the mean achievement of students might vary with institution size and institution type (public vs. private); a stratified sample was used in order to control any such effects, without necessarily presupposing them. It was of interest, therefore, to see whether foreign language achievement was indeed associated with the stratification variables, and if so, in what ways.

Table 5.4 shows the mean MLA skills test scores, for students of French, German, and Spanish classified by institution size and type. "Institution size" was defined as it had been in the sample design itself, i.e. in terms of the total number of foreign language majors graduating in 1963 according to U. S. Office of Education figures. Only "regular" cases with complete sets of MLA skills test scores were employed in the tabulation. The number of students in Russian was too small to warrant analysis according to this scheme. Significance tests are given for the "main effects" in a two-way analysis of variance with unequal numbers in cells: the stratum effect, the institution type effects, and the interaction between stratum and institution type. The stratum and institution type effects concern whether there are overall differences, respectively, among strata and between types. The interaction effect concerns whether similar stratum differences are to be observed in both public and private institutions (or, what amounts to the same thing, whether similar institution type differences are to be observed in all strata). For the stratum and institution type effects, a "significant" effect is to be interpreted as indicating significant differences among strata or between types. For the interaction effect, a significant effect means that differences in one of the variables are not identical in each classification of the other variable.

Institution size (i.e., stratum in the sampling design) is a highly significant effect in all the language groups and for all four of the MLA skills tests. The overall trend is for the students at the larger institutions to obtain higher mean scores. We will discuss the possible reasons for this after reviewing more data.

Institution type is a very highly significant effect in French, a generally highly significant effect in Spanish (the exception being the data for the Speaking test), and a somewhat less significant effect in German. In the German group, public and private institutions do not differ, overall, in mean Listening and Speaking scores, but students at the private institutions do tend to be superior in Reading and Writing. The general trend is for the private institutions to have students with higher average achievement than the public institutions do.

Small but significant interaction effects are to be observed in the data for the German and Spanish groups, but not for the French group. This effect apparently reflects a trend in both German and Spanish whereby the largest (stratum 5) private institutions exhibit higher mean scores, relatively, than they do at public institutions, while the public and private institutions differ only very little in the strata (1 through 4) containing the smaller institutions. Indeed, this interaction effect, in German and Spanish, goes far toward accounting for all the differences between institution types. The result can be understood most easily if it is couched in the following terms: "if you want to find students with high achievement, go to the largest institutions, especially the private ones."

Possible Explanations for Variations among Strata and between Institution Types

Section 1 of this chapter showed that students who volunteered to take the tests tended to be slightly better students, on the average, than those who failed to volunteer for the tests. Since student response rate tended to be smaller at the larger institutions, this fact might account for some of the tendency for larger institutions to exhibit

Table 5.4a

Mean MLA Skills Test Scores, By Stratum and Institution Type

FRENCH

Stratum	Size as of '62-'63[1]	N	MLA Skills Test Scores			
			Listening	Speaking	Reading	Writing
			PUBLIC INSTITUTIONS			
5	40+	203	44.13	75.62	48.60	47.74
4	20-39	67	42.96	74.13	48.21	47.13
3	10-19	61	40.52	69.26	44.70	42.51
2	5-9	51	40.41	70.96	44.69	43.78
1	1-4	27	37.37	67.52	40.37	38.63
Mean of Means, Total		409	41.08	71.50	45.31	43.96
			PRIVATE INSTITUTIONS			
5	40+	111	47.76	77.24	53.32	52.11
4	20-39	188	47.32	75.64	51.86	49.74
3	10-19	198	45.19	73.08	49.87	48.24
2	5-9	111	42.82	71.89	46.94	44.86
1	1-4	84	38.77	68.60	44.85	41.64
Mean of Means, Total		692	44.37	73.29	49.37	47.32
			ALL INSTITUTIONS (Means of Means)			
5	40+	314	45.94	76.43	50.96	49.92
4	20-39	255	45.14	74.88	50.04	48.43
3	10-19	259	42.85	71.17	47.28	45.37
2	5-9	162	41.61	71.42	45.82	44.32
1	1-4	111	38.07	68.06	42.61	40.13
Mean of Means, Total		1101	42.72	72.39	47.34	45.64

SIGNIFICANCE TESTS (F-ratios)

Effect	ndf's	Listening	Speaking	Reading	Writing
Stratum	4, 1091	29.3**	22.3**	21.8**	23.9
Type	1, 1091	58.0**	10.2**	59.2**	35.9**
Stratum x Type	4, 1091	1.1	0.7	0.8	1.6

[1] Total number of FL majors

**$p < .01$ *$p < .05$

Table 5.4b

Mean MLA Skills Test Scores, By Stratum and Institution Type

GERMAN

			MLA Skills Test Scores			
Stratum	Size as of '62-'63[1]	N	Listening	Speaking	Reading	Writing
			PUBLIC INSTITUTIONS			
5	40+	75	44.20	86.77	51.20	53.97
4	20-39	22	42.73	87.41	50.59	49.45
3	10-19	24	45.33	88.62	51.71	53.08
2	5-9	14	37.07	82.86	45.00	43.07
1	1-4					
Mean of Means, Total		135	42.33	86.41	49.62	49.89
			PRIVATE INSTITUTIONS			
5	40+	41	44.44	95.34	58.46	61.93
4	20-39	43	43.19	87.12	50.81	52.09
3	10-19	63	41.46	83.81	49.75	49.49
2	5-9	38	39.89	80.71	45.97	45.55
1	1-4					
Mean of Means, Total		185	43.49	86.74	51.25	52.26
			ALL INSTITUTIONS (Means of Means)			
5	40+	116	46.82	91.06	54.83	57.95
4	20-39	65	42.96	87.26	50.70	50.77
3	10-19	87	43.39	86.22	50.73	51.28
2	5-9	52	38.48	81.78	45.48	44.31
1	1-4					
Mean of Means, Total		320	42.91	86.58	50.44	51.08

SIGNIFICANCE TESTS (F-ratios)

Effect	ndf's				
Stratum	3, 312	9.2**	6.3**	9.4**	11.6**
Type	1, 312	2.5	1.3	5.1	4.1
Stratum x Type	3, 312	4.5**	4.8**	4.1**	3.0*

---

[1] Total number of FL majors

**p < .01 *p < .05

Table 5.4c

Mean MLA Skills Test Scores, By Stratum and Institution Type

SPANISH

			MLA Skills Test Scores			
Stratum	Size as of '62-'63[1]	N	Listening	Speaking	Reading	Writing
			PUBLIC INSTITUTIONS			
5	40+	167	44.86	83.97	47.92	56.51
4	20-39	72	45.85	88.76	48.32	56.94
3	10-19	81	42.79	80.36	46.75	52.00
2	5-9	70	41.50	81.76	45.56	50.61
1	1-4	25	41.24	81.16	44.80	48.28
Mean of Means, Total		415	43.25	83.20	46.67	52.87
			PRIVATE INSTITUTIONS			
5	40+	78	48.68	88.87	53.36	59.36
4	20-39	92	46.24	86.99	49.52	56.33
3	10-19	101	44.23	83.42	47.34	53.95
2	5-9	79	42.02	78.82	44.80	51.18
1	1-4	36	44.11	82.97	46.11	52.92
Mean of Means, Total		386	45.06	84.21	48.23	54.75
			ALL INSTITUTIONS (Means of Means)			
5	40+	245	46.77	86.42	50.64	57.94
4	20-39	164	46.04	87.88	48.92	56.64
3	10-19	182	43.51	81.89	47.04	52.98
2	5-9	149	41.76	80.29	45.18	50.89
1	1-4	61	42.67	82.06	45.45	50.60
Mean of Means, Total		801	44.15	83.71	47.45	53.81

SIGNIFICANCE TESTS (F-ratios)

Effect	ndf's				
Stratum	4, 791	14.3**	11.0**	9.5**	15.4**
Type	1, 791	14.9**	2.3	10.1**	5.1**
Stratum x Type	4, 791	2.3	3.5**	3.9*	1.2

---

[1] Total number of FL majors

**p < .01   *p < .05

higher means on the MLA skills tests. That is, it is conceivable that if all students at all participating institutions could have been tested, the addition of the weaker students at the larger institutions would reduce or even eliminate the institutional size differences we have just demonstrated. There is no sure way of determining to what extent this could have been the case. It has already been noted, however, that the amount of bias introduced by student non-response is probably fairly small. It is not likely that differences in student response rate could account for all the differences among mean MLA skills test scores at institutions of different sizes.

The data collected in this study make it possible to investigate other sources of variation among institutions of different sizes. From responses on the Questionnaire for Foreign Language Majors, a number of variables were created that reflect characteristics of the students and of the kind of instruction to which they had been exposed. Students in the various cross-classifications of institution size and type were compared on a number of these variables, with results as shown in Table 5.5. This table contains only the F-ratios computed from the two-way analyses of variance, and indications of the extent to which these F-ratios were statistically significant, making it possible to identify those variables that appear to differentiate institutions of different sizes and types. (For comparative purposes, F-ratios are also given for the four MLA skills tests. These ratios are slightly different from those presented in Table 5.4 because of slight differences in the number of cases used for this new analysis.)

Chapters VI and VII will discuss the student and instructional variables more fully, but for the time being it will suffice to give brief descriptions of them:

Time Began: Coded, 1, 2, or 3 respectively depending upon whether the student started the study of his major foreign language in "grade school", in high school, or in college.

Time Abroad: Coded 1 for the student who had never been abroad to the country where the foreign language is used, 2 for the student who had toured in that country and/or had a summer course there, and 3 for the student who had had a year of study abroad.

Use FL at Home: The question was "Do you and/or your parents speak your major language at home?" Answers were coded: 1 for "No", 2 for "Yes, occasionally," and 3 for "Yes, frequently."

Hours College Grammar Courses: Total number of course-hours taken in beginning, intermediate, or advanced courses devoted primarily to language study (as contrasted with literature). For convenience these are called "grammar courses" here.

Teacher FL Use(Grammar Courses): Coded 1, 2, or 3 depending on how much the teacher was reported as using the foreign language in the classroom, averaged over "grammar" courses as defined above.

Student FL Use (Grammar Courses): Coded 1 or 2 depending on how much the student reported he was required to speak the foreign language in the classroom, averaged over "grammar" courses.

Language Laboratory Use (Grammar Courses): Coded 1, 2, or 3 depending upon the extent to which use of a language laboratory was reported as being "an important and integral part" of the course, averaged over "grammar" courses.

Teacher's Accent (Grammar Courses): Coded 1, 2, or 3 depending upon the extent to which the teacher was reported as being a native speaker of the FL, or having a "native" accent, averaged over "grammar" courses.

Hours College Literature Courses: Total number of course-hours taken in literature courses at the college level. A literature course was defined in the questionnaire as "devoted primarily to the reading of texts in the foreign language (not in translation),

Table 5.5

F-Ratios for Stratum and Institution-Type Effects for Selected Variables

	French			German			Spanish		
	Stratum	Type	S x T	Stratum	Type	S x T	Stratum	Type	S x T
Variable ndf's	4, 1090	1, 1090	4, 1090	3, 311	1, 311	3, 311	4, 790	1, 790	4, 790
Time Began	1.2	11.5**	1.5	0.4	2.0	0.9	2.5*	2.1	0.6
Time Abroad	6.7**	46.4**	2.7*	1.7	13.4**	7.7*	3.1*	21.6**	1.8
Use FL at home	1.8	0.7	1.0	2.5	2.0	0.1	2.5*	3.9*	1.9
Hrs. Coll. Gram. Courses	8.1**	4.1*	2.3	2.4	1.5	1.0	1.7	11.1**	1.5
Teacher FL Use (Gram.Courses)	6.7**	30.5**	2.0	2.4	7.7**	0.4	5.5**	6.2*	1.2
Student FL Use (Gram.Courses)	4.9**	34.2**	1.2	2.0	4.3*	0.8	4.2**	4.3*	0.8
Lang. Lab. Use (Gram.Courses)	2.5*	0.0	5.7**	5.3**	2.1	3.2*	3.5**	11.4**	2.4*
Teacher's Accent (Gram.Courses)	6.0**	2.8	1.4	0.7	1.7	0.2	2.7*	3.5	2.0
Hrs. College Lit. Courses	7.3**	58.3**	1.3	1.7	2.5	4.6**	4.3**	21.0**	3.4**
Teacher FL Use (Lit. Courses)	5.1**	12.3**	2.5*	1.4	2.4	6.7**	5.2**	1.2	5.2**
Student FL Use (Lit. Courses)	6.2**	20.1**	4.3**	0.7	2.4	4.9**	2.7*	3.6	4.0**
Lang. Lab. Use (Lit. Courses)	12.7**	0.1	7.0**	4.8**	0.4	1.2	15.0**	6.4*	5.5**
Teacher's Accent (Lit. Courses)	2.9**	9.6**	0.6	1.5	0.6	0.7	1.0	0.8	5.2**
Hrs. Misc. Courses	1.0	1.2	0.6	0.5	1.3	2.3	0.7	9.0**	7.5**
MLA Listening Test	29.1**	59.0**	1.2	8.5**	2.7	4.6**	14.5**	15.4**	2.4*
MLA Speaking Test	22.2**	10.5**	0.7	5.9**	1.4	4.8**	11.2**	2.6	3.7**
MLA Reading Test	21.8**	58.9**	0.8	9.0**	5.2*	4.1**	9.8**	10.7**	4.3**
MLA Writing Test	23.8**	36.5**	1.6	11.1**	4.3*	3.0*	15.5**	5.3*	1.3
$F_{.05}$	2.38	3.85	2.38	2.61	3.87	2.61	2.39	3.86	2.39
$F_{.01}$	3.34	6.66	3.34	3.85	6.72	3.85	3.35	6.67	3.35

** p<.01   * p<.05

rather than the study of the language as such."

Teacher FL Use (Literature Courses): Same as Teacher FL Use (Grammar Courses) except that it was averaged over literature courses.

Student FL Use (Literature Courses): Analogous to Student FL Use (Grammar Courses).

Language Laboratory Use (Literature Courses): Analogous to the same variable for Grammar Courses.

Teacher's Accent (Literature Courses): Analogous to the same variable for Grammar Courses.

Hours Miscellaneous Courses: Hours in any foreign language courses not classified by the student as either beginning, intermediate, or advanced language courses, or as literature courses.

It will be seen from Table 5.5 that every one of these variables yielded one or more significant F-ratios for one or more design effects in one or more languages. Stratum differences were particularly pronounced in French, less so in Spanish, and only occasionally in German. Institution type differences occurred frequently in both French and Spanish and also occasionally in German. There were scattered interaction effects of significance in all the languages. In order to analyze the nature of these effects, we present Table 5.6, which gives the means of selected variables within cross-classifications by stratum and institution type. These are the variables that appear to yield significant effects most consistently across languages, or that have other features of interest. Thus, the first two variables were included (even though Time Began shows only one significant F-ratio) because (as will be shown in Chapter VII), both had been found to be strongly associated with foreign language skill attainment.

> Technical Note: The values labeled "Means of Means" in Table 5.6 are just that; they are not means obtained by pooling cases, because such pooling would obscure the contrasts that are tested by the two-way analysis of variance procedure, which assumes that there are equal numbers of cases in each cross-classification and thus that the cases available in each cross-classification are representative of the population that would fall in that cell.

We will discuss the results in Table 5.6 by examining each variable in turn.

Time Began. The overall means indicate that there are more students who started early (in grade school or in high school) in French and Spanish than in German. Only in Spanish are there even barely significant differences among strata with respect to this variable; early starters show a slight tendency to go to larger institutions. In French there is a highly significant difference between public and private institutions. The plain fact seems to be that students who start early in French are much more likely to go to private institutions. This probably reflects the fact that FLES programs (at least those in the middle and late 1950's, when these students were in elementary school) are more likely to be found in the wealthier schools or school districts ("well-heeled suburbs," as they are sometimes called) -- districts that are more likely to send their college-bound students to private institutions. There is a slight trend of the same sort in Spanish, but it is not significant. (From data to be presented later, this may reflect a difference associated with geography. French is more likely to be studied in the Eastern part of the U.S., where there are more private institutions. Spanish FLES is more widely distributed throughout the U.S., and its products are more evenly distributed between private and public institutions.)

Since Time Began is quite strongly associated with foreign language attainment (see Chapter VII), we may infer, at this point, that one of the reasons why French students at private institutions are so superior in foreign language skills to those at public institutions (see Table 5.4) in that on the average they started their foreign language study at an earlier stage.

Table 5.6a

Means of Selected Students and Instructional Variables, By Stratum and Institution Type

FRENCH

		Student Variables		Instructional Variables				
Stratum	N	Time Began	Time Abroad	Teacher's Language (G)	Lang. Lab. (G)	Hrs. Coll. Lit.	Teacher's Lang. (L)	Lang. Lab. (L)
PUBLIC INSTITUTIONS								
5 (Large)	203	2.07	1.60	2.38	1.68	15.5	2.57	1.01
4	66	2.12	1.41	2.41	1.69	17.4	2.56	1.11
3	61	2.11	1.44	2.32	1.70	15.4	2.48	1.36
2	51	2.10	1.59	2.48	1.53	15.6	2.67	1.12
1 (Small)	27	2.04	1.30	2.36	2.02	12.1	2.61	1.12
Mean of Means	408	2.09	1.47	2.39	1.72	15.2	2.58	1.14
PRIVATE INSTITUTIONS								
5 (Large)	111	1.86	1.92	2.69	1.54	20.0	2.67	1.06
4	188	2.01	1.98	2.68	1.78	20.1	2.69	1.17
3	198	1.96	1.89	2.50	1.69	20.8	2.78	1.15
2	111	2.05	1.67	2.64	1.90	18.0	2.79	1.26
1 (Small)	84	2.13	1.45	2.29	1.70	17.5	2.46	1.11
Mean of Means	692	2.00	1.78	2.56	1.72	19.3	2.68	1.15
ALL INSTITUTIONS (Means of Means)								
5 (Large)	314	1.96	1.76	2.54	1.61	17.8	2.62	1.04
4	254	2.06	1.70	2.54	1.74	18.7	2.62	1.14
3	259	2.03	1.66	2.41	1.70	18.1	2.63	1.26
2	162	2.07	1.63	2.56	1.71	16.8	2.73	1.19
1 (Small)	111	2.08	1.38	2.32	1.86	14.8	2.53	1.12
Mean of Means	1100	2.05	1.62	2.48	1.72	17.3	2.63	1.15
SIGNIFICANCE TESTS (F-Ratios)								
Effects	ndf's							
S	4, 1090	1.2	6.7**	6.7**	2.5*	7.3**	5.1**	12.7**
T	1, 1090	11.7**	46.4**	30.5**	0.0	58.3**	12.3**	0.1
S x T	4, 1090	1.5	2.7*	2.0	5.7**	1.3	2.5	7.0**

Table 5.6b

GERMAN

		Student Variables		Instructional Variables				
Stratum	N	Time Began	Time Abroad	Teacher's Language (G)	Lang. Lab. (G)	Hrs. Coll. Lit.	Teacher's Lang. (L)	Lang. Lab. (L)
				PUBLIC INSTITUTIONS				
5 (Large)	75	2.48	1.48	2.01	1.43	14.2	2.29	1.04
4	22	2.68	1.64	2.17	1.54	16.9	2.41	1.01
3	24	2.62	2.21	2.00	1.93	19.9	2.66	1.07
2 }	14	2.57	1.57	1.94	1.38	18.6	2.11	1.07
1 (Small) }								
Mean of Means	135	2.59	1.72	2.03	1.57	17.4	2.37	1.05
				PRIVATE INSTITUTIONS				
5 (Large)	41	2.51	2.29	2.31	1.57	19.9	2.73	1.04
4	43	2.51	1.91	2.39	1.65	16.7	2.53	1.03
3	63	2.41	1.90	2.10	1.73	18.1	2.22	1.04
2 }	37	2.45	1.84	2.09	1.89	16.9	2.36	1.24
1 (Small) }								
Mean of Means	184	2.47	1.98	2.22	1.71	17.9	2.46	1.09
				ALL INSTITUTIONS (Means of Means)				
5 (Large)	116	2.50	1.88	2.16	1.50	17.0	2.51	1.04
4	65	2.60	1.78	2.28	1.60	16.8	2.47	1.02
3	87	2.52	2.06	2.05	1.83	19.0	2.44	1.06
2 }	51	2.51	1.70	2.02	1.63	17.8	2.24	1.16
1 (Small) }								
Mean of Means	319	2.53	1.85	2.13	1.64	17.7	2.41	1.07
				SIGNIFICANCE TESTS (F-Ratios)				
Effects	ndf's							
S	3, 311	0.4	1.7	2.4*	5.3**	1.7	1.4	4.8**
T	1, 311	2.0	13.4**	7.7**	2.1	2.5	2.4	0.4
S x T	3, 311	0.9	7.7**	0.4	3.2*	4.6**	6.7**	1.2

Table 5.6c

SPANISH

		Student Variables		Instructional Variables				
Stratum	N	Time Began	Time Abroad	Teacher's Language (G)	Lang. Lab. (G)	Hrs. Coll. Lit.	Teacher's Lang. (L)	Lang. Lab. (L)
PUBLIC INSTITUTIONS								
5 (Large)	167	2.04	1.66	2.50	1.40	15.50	2.74	1.02
4	72	2.08	1.64	2.44	1.40	19.69	2.81	1.05
3	81	2.17	1.47	2.29	1.55	16.97	2.46	1.02
2	70	2.19	1.57	2.24	1.46	14.71	2.53	1.12
1 (Small)	25	2.08	1.52	2.02	1.62	14.17	2.25	1.20
Mean of Means	415	2.11	1.57	2.30	1.49	16.21	2.56	1.08
PRIVATE INSTITUTIONS								
5 (Large)	77	1.96	2.08	2.48	1.46	20.20	2.63	1.01
4	92	2.00	1.84	2.54	1.74	19.15	2.69	1.04
3	101	2.08	1.71	2.50	1.62	18.68	2.61	1.12
2	79	2.14	1.59	2.36	1.72	19.34	2.65	1.35
1 (Small)	36	2.22	1.72	2.29	1.54	15.67	2.72	1.10
Mean of Means	385	2.08	1.79	2.43	1.62	18.61	2.66	1.12
ALL INSTITUTIONS (Means of Means)								
5 (Large)	244	2.00	1.87	2.49	1.43	17.85	2.68	1.02
4	164	2.04	1.74	2.49	1.57	19.42	2.75	1.04
3	182	2.12	1.59	2.40	1.58	17.82	2.53	1.07
2	149	2.16	1.58	2.30	1.59	17.02	2.59	1.23
1 (Small)	61	2.15	1.62	2.16	1.58	14.92	2.48	1.15
Mean of Means	800	2.10	1.68	2.37	1.55	17.41	2.61	1.10
SIGNIFICANCE TESTS (F-Ratios)								
Effects	ndf's							
S	4, 790	2.5*	3.1*	5.5**	3.5**	4.3**	5.2**	15.0**
T	1, 790	2.1	21.6**	6.2*	11.4**	21.0**	1.2	6.4*
S x T	4. 790	0.6	1.8	1.2	2.4*	3.4**	5.2**	5.5**

Time Abroad. This variable yielded highly significant differences in institution types in all three languages; students at private institutions have on the average more experience in travel or study abroad. Since this variable is also known to be very strongly associated with foreign language skill attainment (see Chapter VII), it constitutes another basis for inferring that students at private institutions show better foreign language skill attainment because they are more likely to have studied or travelled abroad.

In French and Spanish, there is a significant stratum difference whereby students at the larger institutions, particularly the private ones, are more likely to have had experience abroad. In German, no overall stratum effect exists, but the interaction effect indicates that in private institutions the travelled students tend to be at the larger institutions while at public institutions they tend to concentrate in medium-sized ones. Again, these differences are likely to be associated with economic factors: it costs more money to go abroad to travel or study. The stratum differences in French and Spanish possibly reflect a greater likelihood that the larger institutions have programs for study abroad.

Although in Table 5.6 Time Began and Time Abroad are labeled as "student variables," this is only because they pertain to possible experiences of the student before he reaches the college level. Actually, they are instructional variables, of a sort, and the Time Abroad variable is most likely to reflect experience during the college years.

We turn now to instructional variables that clearly pertain solely to experiences during the college years.

Teacher FL Use (Grammar Courses). This variable yields significant institution type differences in all languages. Teachers in private institutions are more likely than those in public institutions to use the foreign language in the classroom, at least for the courses actually taken by the students in the sample. (The fact that the mean for this variable is particularly high in French may reflect the fact that the French students are more likely to be advanced students by the time they arrive at college.) There are also stratum differences in all languages, but in French these differences are apparently not perfectly correlated with the size of the institution. There tends to be most use of French in the classroom in stratum 2, the next to smallest; after this come strata 4 and 5, the two largest strata, then stratum 3, and finally stratum 1, the smallest (see the means of means given for "all institutions"). On the other hand, in Spanish (and also in German) there is a trend whereby the foreign language is more often used in the classroom in the larger institutions than in the smaller institutions. Insofar as the teacher's use of the foreign language in the classroom may be a factor in the student's attainment of foreign language skill, this is another possible cause of differences in the mean skill attainment shown by students at different types and sizes of institutions.

Language Laboratory Use (Grammar Courses). The relationships exhibited by this variable are very complex, and seem to be different across languages. In French, there is no overall difference between private and public institutions, but there are significant stratum and stratum x type effects whereby the smaller institutions, particularly the public ones, are more likely to use language laboratories as an integral part of a language course. Possibly in this way they attempt to substitute for the lack of teachers with a native accent. A somewhat similar phenomenon occurs in German, except that strikingly high use of the language laboratory occurs in public institutions of medium size (stratum 3) and in private institutions of small size (strata 1 and 2). In Spanish the major effect is associated with institution type, although there are also stratum and interaction effects: private institutions, particularly the smaller ones, tend to use language laboratories more often.

Hours of College Literature Courses. Here again the differences are of a complex order and vary from language to language. In French and Spanish both stratum and institution type differences are found: students at private institutions nearly always report more hours in literature courses, and the average number of hours increases with the size of the institution. In German, on the other hand, the only effect is an interaction one whereby more hours are reported in smaller public institutions and larger private institutions.

Teacher FL Use (Literature Courses). In French, private institutions show higher overall averages in this variable, but the trends by strata are reversed; that is, in private institutions the averages are positively correlated with institution size but in public institutions they are negatively correlated with institution size. A similar interaction effect occurs in German and Spanish. In Spanish there is a significant stratum effect (overall, larger institutions are higher on this variable), but no overall institution type difference.

Language Laboratory Use (Literature Courses). The overall magnitude of the means (in the range near 1) indicates that there is seldom any use of the language laboratory in literature courses. Where it occurs, it is most likely to occur in the smaller or medium-small institutions, and this effect is highly significant in all languages.

In view of these findings, it is not surprising that institutions of different sizes and types vary considerably in the mean attainment of these students in foreign language skills. Our best guess is that even if there had been a uniformly high student response rate in all institutions, the various strata and institution types would still have shown highly significant differences in foreign language test scores.

As one test of this possibility, a special analysis of institutional differences in FL skill attainment was made with a statistical control for two student variables that are strongly associated with FL skill attainment: Time Began and Time Abroad. That is, the cross-classification means for the MLA skills test scores were adjusted for any effects that might have been associated with differences in the time students began study of their major foreign language or in the amount of travelling and studying abroad they had done. The results are shown in Table 5.7. The adjustments tended to reduce the size of the F-ratios, particularly those associated with interaction effects. As expected, however, both stratum and institution-type differences tended to remain significant, with the same trends as observed before. The effect of stratum was most resistant to the effects of the adjustment; in fact, in German the stratum difference was enhanced by the adjustment. Institution-type differences were virtually eliminated in the German and Spanish groups but remained significant (although not nearly as strongly) in French, except in the case of the Speaking test, where the difference disappeared after the adjustment. The conclusion suggested by these findings is that larger institutions, and to some extent the private ones, produce better language students than smaller ones, even after student backgrounds are taken into account.

3. Variations in Foreign Language Skill Associated with Geographical Region

Geographical region of the U. S. was not explicitly used as a basis for the sampling design, since institutions were sampled at random from all parts of the country. It is of interest, nevertheless, to investigate any variations in foreign language skill attainments that might be associated with differences in geographical region of the institutions sampled. For this purpose, all "regular" cases in French, German, and Spanish with complete information on MLA skills test scores were classified as to which of eight geographical regions their institutions were located in. These geographical regions are those used in certain tabulations of the U. S. Office of Education, as follows:

1. New England: Conn., Maine, Mass., N. H., R. I., Vt.
2. Mideast: Del., D. C., Md., N. J., N. Y., Penna.
3. Great Lakes: Ill., Ind., Mich., Ohio, Wisc.
4. Plains: Iowa, Kans., Minn., Missouri, Neb., N. D., S. D.
5. Southeast: Ala., Ark., Fla., Ga., Ky., La., Miss., N. C., S. C., Tenn., Va., W. Va.

Table 5.7

F-Ratios for Stratum and Institution-Type Effects for MLA Skills Test Scores, Without and With Adjustment for Variations in the Time the Student Began FL Study and in Amount of Travel and Study Abroad

MLA Test	ndf's[1]	French			German			Spanish		
		Stratum 4, 1088	Type 1, 1088	S x T 4, 1088	Stratum 3, 309	Type 1, 309	S x T 3, 309	Stratum 4, 788	Type 1, 788	S x T 4, 788
Listening	Unadj.'d	29.1**	59.0**	1.2	8.5**	2.7	4.6**	14.5**	15.4**	2.4*
	Adjusted	23.1**	20.0**	0.2	17.3**	0.8	0.6	10.2**	4.4*	1.7
Speaking	Unadj.'d	22.2**	10.5**	0.7	5.9**	1.4	4.8**	11.2**	2.6	3.7**
	Adjusted	19.5**	0.0	0.7	10.6**	1.0	1.1	8.3**	0.0	2.7*
Reading	Unadj.'d	21.8**	58.9**	0.8	9.0**	5.2*	4.1**	9.8**	10.7**	4.3**
	Adjusted	17.0**	29.8**	0.7	14.3**	0.1	0.7	6.4**	2.7	3.1*
Writing	Unadj.'d	23.8**	36.5**	1.6	11.1**	4.3*	3.0*	15.5**	5.3*	1.3
	Adjusted	19.0**	11.7**	1.3	18.1**	0.0	0.1	11.4**	0.4	1.2

[1] No. of degrees of freedom for Adjusted F-ratios. See Table 5.5 for Unadjusted F-ratios.

**$p < .01$ *$p < .05$

6. Southwest: Ariz., N. M., Okla., Tex.
7. Rocky Mountain: Colo., Idaho, Mont., Utah, Wyo.
8. Far West and Pacific: Alaska, Calif., Hawaii, Nev., Ore., Wash.

Table 5.8 shows the percentages of French, German, and Spanish majors found in each of these eight regions, and the mean scores of these majors on each of the MLA skills tests in the relevant language.

The percentages reflect in part, of course, the geographical distribution of institutions over these regions. For example, there are large numbers of institutions in the Mideast region, and many of these institutions were included in the national study sample. There are interesting variations, however, in the distribution of FL majors over the regions. Examining only percentages (that is, laying aside the fact that French is overall the most popular language, Spanish the next most popular, and German a trailing third), we find that French is relatively more popular than Spanish or German in the New England and Mideast states. Spanish is relatively more popular in the Plains, Southeast, and Southwest states, and German relatively more popular in the Great Lakes, Rocky Mountain, and Far West and Pacific states. These differences correspond, roughly, to the known orientations of the different regions towards the different languages. The Northeast has traditionally had closer ties with France and (to some extent) French Canada; the Southeast and Southwest have strong Spanish influences in their histories and contemporary lives; the Great Lakes states were those where large numbers of German immigrants went to settle.

Regional differences in MLA skills test scores are in every case statistically significant, highly so in French and Spanish where results are available for large numbers of students. Generally, the means are higher in those regions where the languages are relatively more popular. At the same time, regions appear to differ in their overall performance, regardless of language. For example, students in New England institutions are generally high in all languages; only in Spanish are they slightly less superior than students in some other regions. Students in the Plains, Southeast, and Southwest states are generally poorer than students from other regions, except that students of Spanish in the Southwest do quite well.

It should be remembered that these regional differences are probably associated in a complex way with other variables mentioned in this chapter: institution size, institution type, the student's time of beginning the foreign language, the student's experiences in study or travel abroad, etc. No attempt has been made in the present report to study regional differences adjusted for the effects of these other variables.

4. Estimated National Norms for MLA Skills Tests

Chapter IV, Section 3 presented means, standard deviations, and percentile norms derived directly from data obtained from the students actually tested. Because the stratified sampling design represented institutions of different sizes at different rates, and because student response rate itself varied over the cells of the sampling design, it was thought desirable to attempt to estimate distributions of test scores for the populations which had been sampled. This would result in national norms that should be good estimates of the norms that would have been obtained had all students in the defined population been tested in the spring of 1965.

Consultations were held with Professor W. G. Cochran of the Harvard University Department of Statistics and with Dr. Theodore Colton of the Harvard Medical School as to the optimal estimation procedures to be used. After considerable study of the problem, it was decided not to make any attempt to correct for the effects of student non-response upon test score distributions, since information as to the amount of such effects was sparse, and since the meager information that did exist (see Section 1 of this chapter) indicated that the effects might be relatively small.

Table 5.8

Mean MLA Skills Test Scores by Geographical Region
(Regular cases)

	N	%	Listening	Speaking	Reading	Writing
			FRENCH			
1. New England	155	13.9	46.32	74.86	51.57	50.12
2. Mid East	335	30.0	44.67	74.55	49.41	48.11
3. Great Lakes	255	22.9	44.19	73.58	47.90	46.57
4. Plains	83	7.4	41.76	71.83	47.13	43.71
5. Southeast	148	13.3	41.35	70.40	46.78	44.60
6. Southwest	25	2.2	40.64	71.16	46.40	43.52
7. Rocky Mountain	34	3.0	46.00	75.74	49.29	47.56
8. Far West & Pacific	80	7.2	45.14	74.25	49.80	48.55
Total	1115	100.0	44.12	73.56	48.81	47.16
p			< .001	< .001	< .001	< .001
			GERMAN			
1. New England	36	11.1	45.44	88.39	54.11	54.67
2. Mid East	64	19.8	43.05	86.70	50.22	51.64
3. Great Lakes	82	25.3	44.34	87.80	51.76	53.30
4. Plains	33	10.2	39.76	82.03	48.24	47.42
5. Southeast	31	9.6	40.13	84.39	48.61	47.94
6. Southwest	7	2.2	41.71	84.29	46.43	48.71
7. Rocky Mountain	29	9.0	45.48	91.28	52.76	56.71
8. Far West & Pacific	42	13.0	44.76	86.10	51.74	52.64
Total	324	100.0	43.44	86.75	51.03	52.09
p			< .05	> .05	> .05	> .05
			SPANISH			
1. New England	61	7.5	46.39	85.67	51.00	57.54
2. Mid East	219	26.9	44.09	80.65	47.37	54.77
3. Great Lakes	134	16.4	43.60	83.80	46.25	53.77
4. Plains	102	12.5	41.81	80.75	44.19	51.16
5. Southeast	109	13.4	42.98	83.64	46.65	51.94
6. Southwest	70	8.6	45.79	88.87	49.86	57.34
7. Rocky Mountain	45	5.5	47.73	92.04	50.89	57.80
8. Far West & Pacific	75	9.2	48.11	88.25	52.05	57.55
Total	815	100.0	44.47	83.99	47.80	54.60
p			< .001	< .001	< .001	< .001

The estimation procedure finally decided upon consisted of two adjustments for effective sampling fractions: (1) the ratio of the number (B) of students actually tested in a particular language, stratum, and institution type, to the number (C) of all students that according to U. S. Office of Education figures were available to be tested in that language, stratum, and institution type in the participating institutions; and (2) the ratio of the number (D) of participating institutions in a given stratum and institution classification to the number (E) of all institutions offering foreign language programs that were in that stratum and institution classification according to U. S. Office of Education figures for 1962-63 (Wright, 1965). The reciprocal of the first ratio, C/B, made a correction for student response rate within a given language, stratum, and institution type classification; the reciprocal of the second ratio, E/D, was the correction for the effective sampling fraction in the sampling design, i. e., it corrected for the fact that the sampling design represented smaller-sized institutions much less than larger-sized institutions.

The ratios thus determined are shown in Table 5.9. The reciprocals of the ratios were applied to the score frequencies for a given language and test to estimate the score frequency for the national population in that language, according to the formula

$$f_{in} = f_{is} \left(\frac{C}{B}\right)\left(\frac{E}{D}\right) ,$$

where f_{in} is the frequency of scores in the ith score interval in the estimated national frequency distribution in a given language and test, and f_{is} is the frequency of scores in the ith score interval in the sample actually tested in a given language and test. The ratio C/B was of course different for each test, language, stratum, and institution type; the ratio E/D was varied only by stratum and institution type; i.e., it was uniform for all languages, on the assumption that the data available in a given stratum and institution classification should be boosted equally for all languages.

It should be noted that the number B, and the values f_{is} were all from data on "regular cases" as defined previously. That is, data from native speakers and other "odd" types of cases with unusual language experiences had been excluded; the estimated national norms are therefore for a population composed exclusively of "regular cases."

An example may help in understanding this procedure. There was 1 student in Public Stratum 1 who obtained a score of 54 on the French Listening test, and there were 2 students who obtained this score in Private Stratum 1. Now, in Public Stratum 1 there were in all 29 students tested (B) out of a total of 31 (C) available to be tested in the participating institutions of that stratum. There were 11 (D) such institutions out of a total of 75 (E) public institutions in that stratum. The ratio (B/C) was .9355, and the ratio (D/E) was .1467, as may be found in Table 5.9. By the formula given above, we may then estimate the total number of students in Public Stratum 1 who would get a score of 54 as 1/[(.9355)(.1467)] = 7.2884. Similarly, in Private Stratum 1 there were in all 87 students tested (B) out of a total of 97 (C) available to be tested in the participating institutions of that stratum. There were 32 (D) such institutions out of a total of 251 (E) private institutions in that stratum. The ratio (B/C) was .8969, and the ratio (D/E) was .1275. We then estimate the total number of students in Private Stratum 1 who would get a score of 54 as 2/[(.8969)(.1275)] = 17.4908. Adding the estimates for public and private institutions, we have 7.2884 + 17.4908 = 24.7792 as the estimated number of students in stratum 1 of the national population who would obtain a score of 54 in the French Listening test.

Summing these and similar computations for the remaining strata and for all possible scores yielded the estimated frequency distribution for the French Listening test in the national population. The estimated number of French students in the national population was the sum of frequencies, 5564.63. The mean and standard deviation of the estimated score distribution were computed as 43.0766 and 7.8700, respectively. (The estimated number of students in the three languages are roughly comparable to the actual numbers in the class of 1965; see Chapter III.)

Table 5.9

Ratios Used in Estimation of National Norms

A. Ratios B/C (see text), by language, test, stratum, and institution type

	PUBLIC INSTITUTIONS				PRIVATE INSTITUTIONS			
	TEST				TEST			
Stratum	L	S	R	W	L	S	R	W
				FRENCH				
5 (large)	.294	.271	.296	.296	.367	.358	.361	.364
4	.449	.376	.449	.449	.523	.512	.542	.539
3	.477	.462	.500	.500	.707	.677	.718	.718
2	.724	.684	.711	.711	.818	.790	.811	.811
1 (small)	.935	.903	.935	.903	.897	.866	.897	.897
				GERMAN				
5 (large)	.236	.227	.233	.230	.459	.429	.449	.459
4	.250	.229	.250	.250	.512	.512	.536	.536
3	.429	.429	.429	.429	.759	.736	.747	.759
2	.333	.333	.333	.333	.543	.543	.543	.543
1 (small)	.500	.500	.500	.500	.833	.792	.833	.833
				RUSSIAN				
5 (large)	.206	.206	.206	.206	.271	.271	.271	.257
4	.300	.300	.300	.250	.429	.405	.429	.429
3	.500	.500	.500	.500	.250	.250	.250	.250
2	--	--	--	--	2.000	2.000	2.000	2.000
1 (small)	1.000	1.000	1.000	1.000	--	--	--	--
				SPANISH				
5 (large)	.252	.246	.252	.257	.410	.405	.405	.410
4	.442	.354	.437	.437	.466	.442	.486	.486
3	.500	.494	.512	.512	.582	.560	.571	.577
2	.809	.787	.809	.798	.934	.912	.934	.934
1 (small)	.619	.595	.619	.619	.720	.720	.720	.720

B. Ratios D/E, by stratum and institution type

Stratum	PUBLIC INSTITUTIONS	PRIVATE INSTITUTIONS
5 (large)	.8000	.9091
4	.4815	.6047
3	.3778	.3441
2	.1803	.2239
1 (small)	.1467	.1275

Table 5.10 shows the N's, means, and standard deviations estimated for all the MLA skills tests. Table 5.11 gives the percentile norms for each test. Each percentile rank is computed for the mid-point of the corresponding score interval. Table 12 gives the estimated percentile norms in a more convenient form than Table 11; they are analogous to the norms presented by Educational Testing Service (1964b) for NDEA Institute teachers (see Appendix E-3). Figures 5.1, 5.2, 5.3, and 5.4 give the norms in graphic form, plotted on scales that are set up to correspond with the FSI "absolute language proficiency" scales; these charts also show the ranges for the Myers-Melton (1964) teacher qualification standards (Superior, Good, Minimal, and Unsatisfactory). For a discussion of the FSI and Myers-Melton equivalences, see Chapter II, Section 2, of this report.

The arrows depicted close to the names of the four skills at the top of each figure will remind the reader that the Listening and Speaking scales were equated to the FSI "S" (Speaking) ratings, while the Reading and Writing scales were equated to the FSI "R" (Reading) ratings. The scale for each MLA skill test is labeled with the so-called "converted" scores, and superimposed on the scale are bars and other indications showing the median, the middle 50 per cent of the cases (from the 25th to the 75th percentile) and the middle 80 per cent of the cases (i.e., from the 10th to the 90th percentile).

The most striking thing about these charts is the generally low median levels of attainment in audio-lingual skills that they reveal. The median graduate with a foreign language major can speak and comprehend the language only at about an FSI Speaking rating of "2+," that is, somewhere between a "limited working proficiency" and a "minimum professional proficiency." There is some variation around this value for the different languages, which the reader may perceive by inspecting the charts. In Reading and Writing, the French, German, and Spanish groups attain median scores that correspond approximately to an FSI rating of "3," that is, "minimum professional proficiency." The slightly poorer showing of the Russian group in all four skills may reflect either the greater difficulty of the Russian language for English speakers or the relatively smaller investment that has been made in the teaching of Russian in the United States, or both. But when we consider all the results, the net showing of college foreign language majors is not impressive. One would think that the median college foreign language major ought to attain more than a "minimum professional proficiency" in basic language skills as a result of his efforts. We are speaking, of course, of medians; the distributions range from abysmally poor performance to quite superior performance--up to the limits measured by the tests, in some cases.

Comparison of CFLT Norms with NDEA Institute Norms

For some years--that is, since about 1963, interpretations of scores on the MLA Advanced Proficiency Tests have necessarily been based on the only norms available--the norms from teachers enrolled at NDEA Institutes in 1961, 1962, and 1963. (These norms may be found in Appendix E-3.) From the point of view of teacher certification and for other purposes, the College Foreign Language Testing (CFLT) norms presented in Tables 11 and 12 will probably be more useful and pertinent than the NDEA norms. It is well, therefore, to recognize the fairly considerable differences between the CFLT and the NDEA norms. Table 5.13 presents a statistical summary of the differences in terms of means and standard deviations. The CFLT norms are, of course, based on smaller actual numbers of cases tested, but for practical purposes they may be considered good estimates of the populations that were sampled, these populations being (by chance) approximately of the same magnitude, numerically, as the samples of teachers at NDEA Institutes in 1961-63.

As a general statement to summarize the differences between the CFLT and the NDEA norms, it may be said that the CFLT samples show clearly higher means than the NDEA pretest norms, and slightly higher means, in general, than even the NDEA *post*-test norms; further, they show much less range or variance than the NDEA norms. Despite what would appear to be considerable range of ability in the CFLT population, it is still smaller *in every comparison* than that shown by the NDEA norms. The CFLT group, it must be observed, could be said to have a *relatively* homogeneous background in the sense that every member has been recently through the experience of preparing himself as a candidate for an A.B.

Figure 5.1

Estimated National Norms for CFLT Population,
with MLA-FSI and Myers-Melton Equivalents

FRENCH

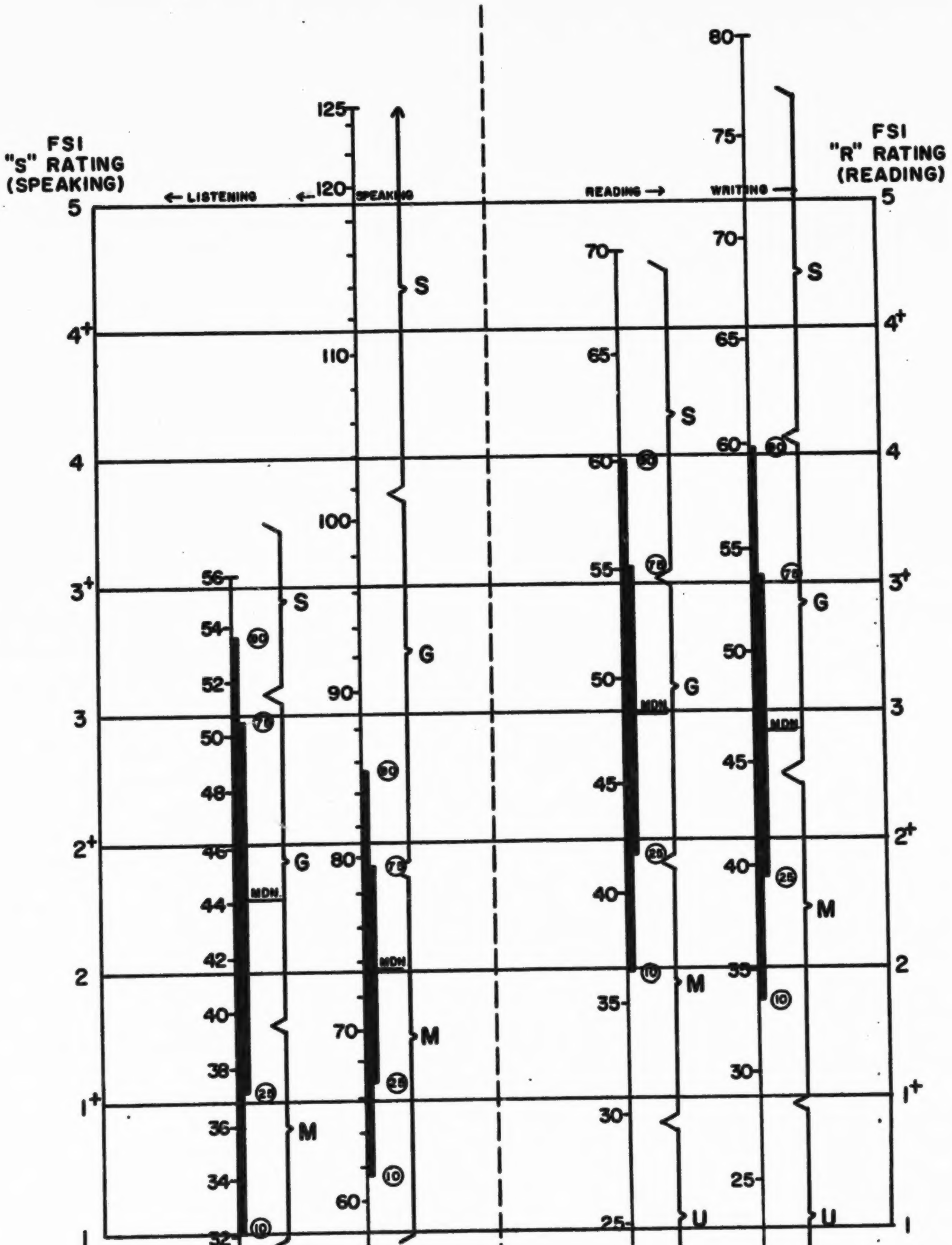


Figure 5.2

Estimated National Norms for CFLT Population, with MLA-FSI and Myers-Melton Equivalents

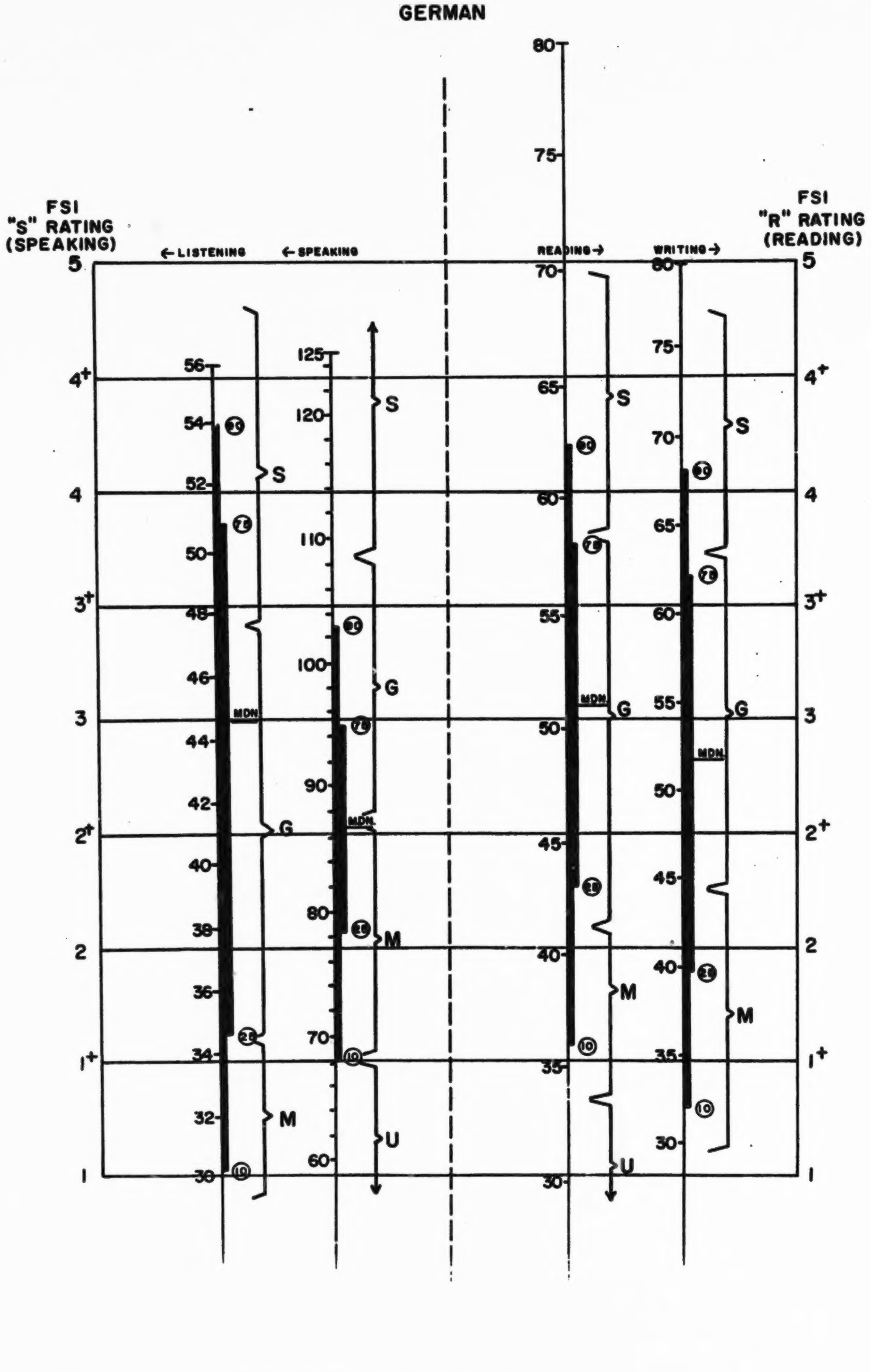


Figure 5.3

Estimated National Norms for CFLT Population,
with MLA-FSI and Myers-Melton Equivalents

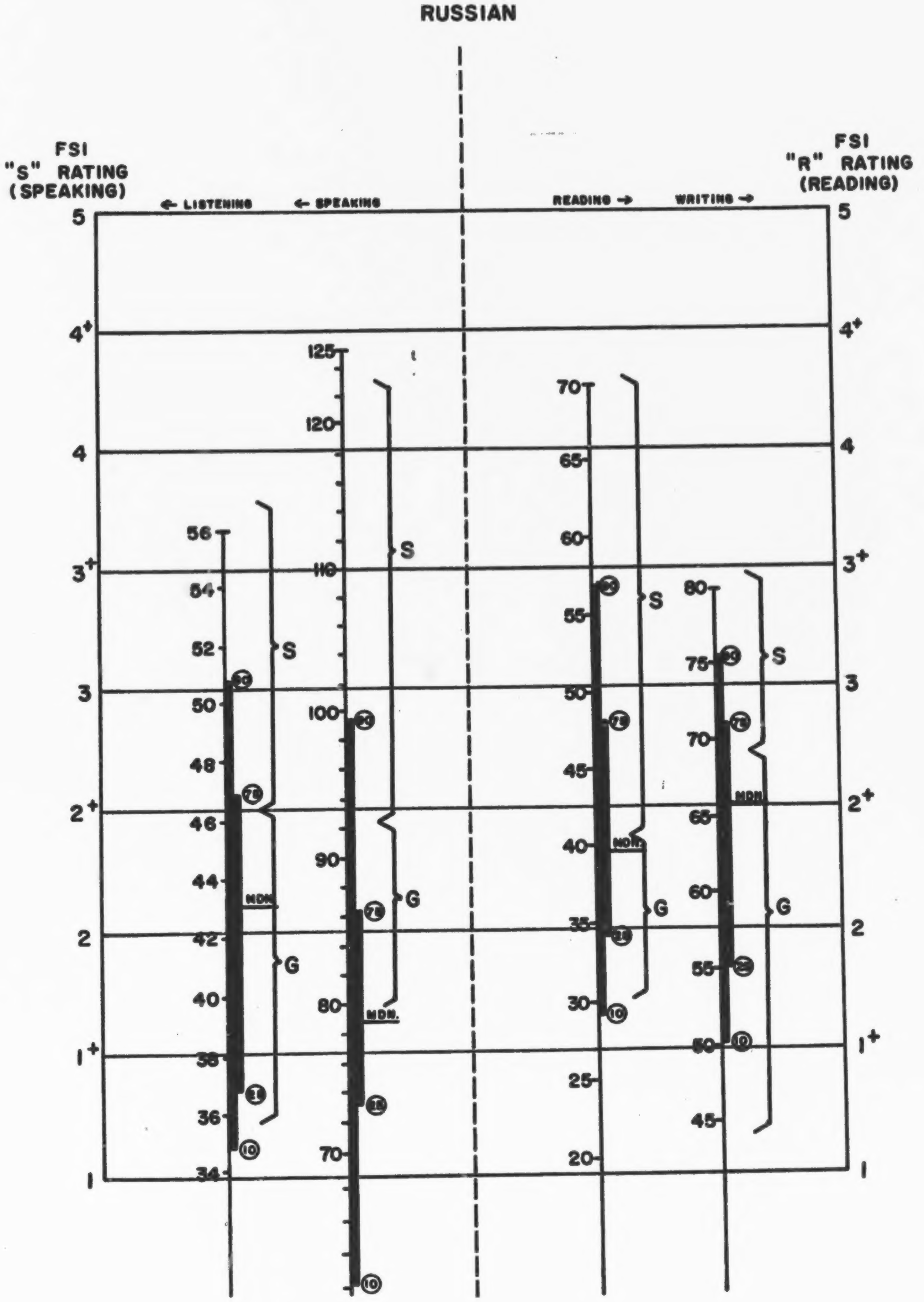


Figure 5.4

Estimated National Norms for CFLT Population, with MLA-FSI and Myers-Melton Equivalents

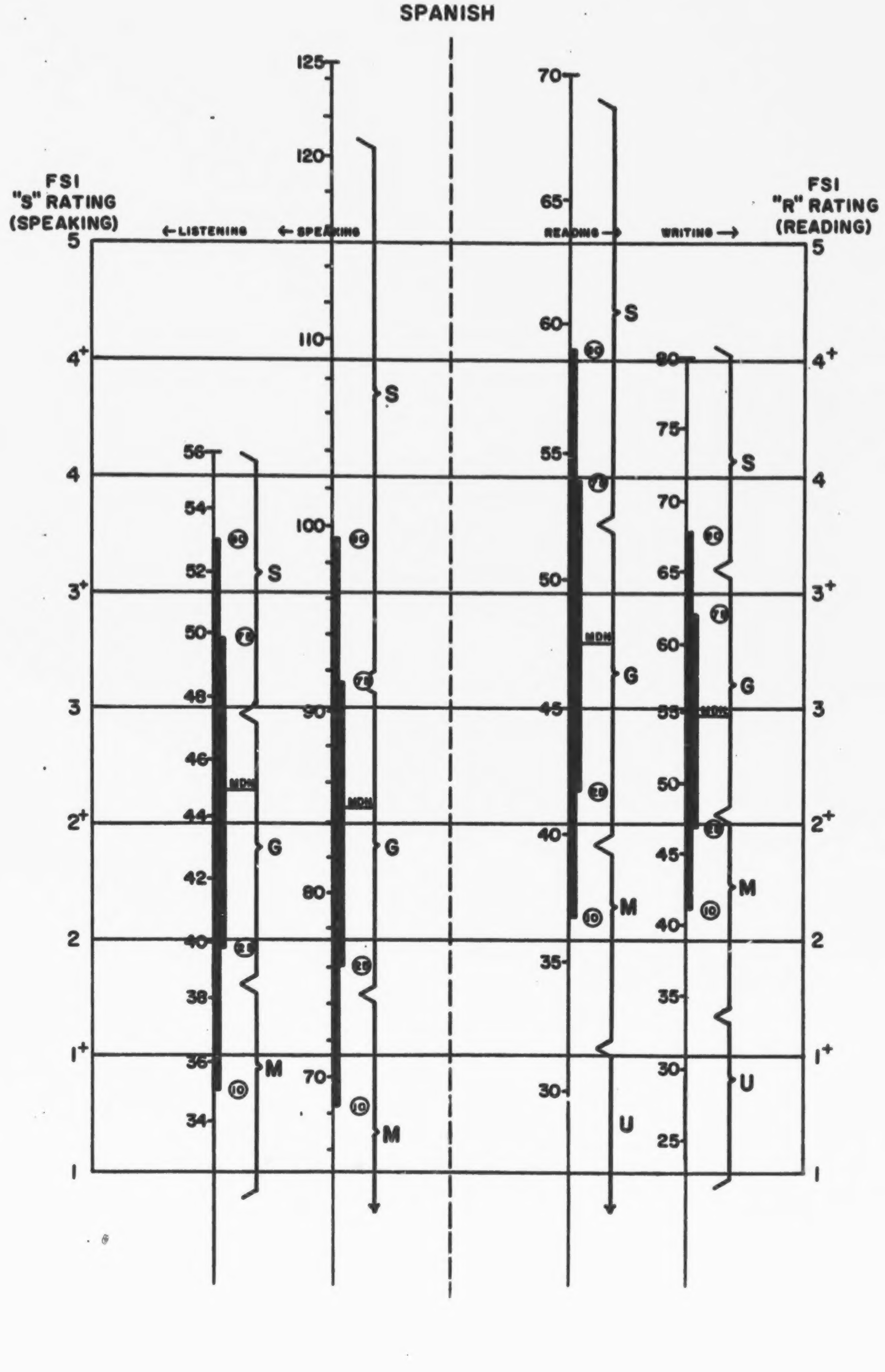


Table 5.10

Means and Standard Deviations for Estimated National Norms

			MLA Skills Test			
Language	Est. N		Listening	Speaking	Reading	Writing
FRENCH	5564.63	$\bar{X}$	43.0766	72.7511	47.8634	46.0483
		σ	7.8700	9.4849	8.8337	9.7730
GERMAN	1847.69	$\bar{X}$	42.5462	85.7506	49.8498	50.4167
		σ	8.7640	12.9107	9.7752	13.5069
RUSSIAN	480.54	$\bar{X}$	42.5283	79.2384	40.9072	62.5431
		σ	6.0294	13.5407	9.6871	10.9354
SPANISH	4439.23	$\bar{X}$	43.9717	83.5241	47.2145	53.8009
		σ	6.7925	12.0231	8.3135	10.1560

Table 5.11

Estimated National Norms for MLA Proficiency Test Skill Scores, College Senior FL Majors (Percentile Ranks of Converted Scores)

FRENCH

Listening		Speaking		Reading		Writing	
Score	%ile Rank	Score	%ile Rank	Score	%ile Rank	Score	%ile Rank
56	99.3	104	99.9	69	100.0	70	99.8
55	97.2	103	99.8	68	99.9	69	99.4
54	93.8	102	99.7	67	99.8	68	99.0
53	89.5	101	99.6	66	99.4	67	98.8
52	85.5	100	99.6	65	98.8	66	98.5
51	80.7	99	99.5	64	98.1	65	98.1
50	75.6	98	99.4	63	96.7	64	97.4
49	71.2	97	99.2	62	95.1	63	96.4
48	66.9	96	99.0	61	93.4	62	95.0
47	62.6	95	98.8	60	91.1	61	93.4
46	58.4	94	98.7	59	88.3	60	91.5
45	54.5	93	98.3	58	85.4	59	89.5
44	51.1	92	97.9	57	82.6	58	87.4
43	46.9	91	97.3	56	79.3	57	85.2
42	42.6	90	96.5	55	76.2	56	83.0
41	39.1	89	95.8	54	73.0	55	80.4
40	35.7	88	94.8	53	69.3	54	77.8
39	32.1	87	93.4	52	65.2	53	74.9
38	29.0	86	92.1	51	61.1	52	71.5
37	25.9	85	90.7	50	57.2	51	67.8
36	22.3	84	89.0	49	53.7	50	63.7
35	19.0	83	87.1	48	49.9	49	60.1
34	16.3	82	84.6	47	45.6	48	57.1
33	13.9	81	82.0	46	41.7	47	54.0
32	11.3	80	79.1	45	38.0	46	50.3
31	8.3	79	75.5	44	34.4	45	46.7
30	5.7	78	71.6	43	31.0	44	43.3
29	3.9	77	67.2	42	27.6	43	39.3
28	2.5	76	62.6	41	24.5	42	35.5
27	1.2	75	58.5	40	21.2	41	32.0
26	0.4	74	54.3	39	18.3	40	28.2
25	0.2	73	49.9	38	15.8	39	24.9
24	0.1	72	45.9	37	13.2	38	21.7
		71	42.1	36	10.9	37	18.7
		70	38.0	35	8.7	36	16.1
		69	34.3	34	6.4	35	14.0
		68	31.1	33	4.5	34	12.5
		67	27.5	32	3.1	33	10.5
		66	23.8	31	2.0	32	8.3
		65	20.2	30	1.5	31	6.5
		64	17.5	29	1.1	30	5.0
		63	15.2	28	0.6	29	3.9
		62	12.6	27	0.1	28	3.1
		61	10.3			27	2.2
		60	8.8			26	1.4
		59	7.4			25	0.7
		58	6.0			24	0.4
		57	4.9			23	0.2
		56	3.7			22	0.1
		55	3.0				
		54	2.5				
		53	1.9				
		52	1.4				
		51	1.1				
		50	0.9				
		49	0.7				
		48	0.6				
		47	0.5				
		46	0.5				
		45	0.4				
		44	0.3				
		43	0.2				
		42	0.2				
		41	0.2				
		40	0.2				
		39	0.1				

Table 5.11, continued

Estimated National Norms for MLA Proficiency Test Skill Scores, College Senior FL Majors (Percentile Ranks of Converted Scores)

GERMAN							
Listening		Speaking		Reading		Writing	
Score	%ile Rank	Score	%ile Rank	Score	%ile Rank	Score	%ile Rank
56	98.8	118	99.9	70	99.9	77	99.9
55	96.0	117	99.9	69	99.7	76	99.6
54	92.3	116	99.9	68	99.1	75	99.3
53	88.6	115	99.8	67	97.3	74	98.7
52	83.1	114	99.7	66	95.5	73	97.7
51	77.3	113	99.3	65	94.0	72	96.6
50	73.3	112	98.8	64	92.5	71	95.5
49	69.0	111	98.4	63	91.6	70	94.1
48	64.7	110	97.6	62	90.0	69	92.5
47	61.7	109	96.7	61	87.8	68	90.7
46	58.5	108	96.1	60	85.3	67	88.8
45	53.6	107	95.4	59	80.6	66	87.1
44	49.7	106	94.6	58	76.4	65	85.4
43	45.8	105	94.1	57	73.7	64	82.3
42	43.0	104	93.3	56	69.7	63	79.2
41	41.5	103	91.4	55	65.2	62	76.5
40	39.8	102	88.2	54	61.3	61	73.0
39	37.3	101	85.6	53	58.5	60	69.7
38	34.7	100	84.9	52	55.6	59	66.8
37	31.8	99	84.4	51	52.1	58	64.1
36	29.2	98	83.0	50	48.3	57	61.3
35	28.0	97	81.6	49	44.3	56	58.2
34	24.5	96	79.6	48	40.4	55	56.6
33	19.4	95	76.2	47	37.6	54	55.1
32	15.9	94	73.4	46	36.0	53	53.1
31	13.1	93	70.5	45	33.6	52	51.2
30	10.5	92	67.3	44	30.0	51	49.8
29	7.7	91	65.1	43	26.3	50	48.2
28	4.1	90	62.7	42	22.9	49	46.3
27	1.3	89	59.3	41	20.0	48	44.5
26	0.6	88	55.3	40	18.1	47	42.5
25	0.5	87	51.9	39	15.6	46	40.1
24	0.2	86	48.8	38	13.5	45	37.2
		85	45.5	37	12.2	44	34.4
		84	42.2	36	10.9	43	31.5
		83	39.5	35	8.6	42	28.6
		82	37.4	34	6.8	41	27.0
		81	34.9	33	5.4	40	25.8
		80	31.9	32	3.3	39	24.2
		79	28.3	31	2.0	38	21.8
		78	25.2	30	1.2	37	19.7
		77	23.8	29	0.7	36	17.6
		76	22.8	28	0.5	35	15.2
		75	21.2	27	0.5	34	13.9
		74	18.8	26	0.5	33	12.5
		73	17.0	25	0.3	32	11.0
		72	15.8			31	9.4
		71	14.1			30	7.8
		70	12.6			29	6.5
		69	11.4			28	5.3
		68	10.3			27	4.9
		67	7.8			26	4.2
		66	5.5			25	2.7
		65	4.6			24	1.3
		64	4.1			23	0.6
		63	3.6			22	0.5
		62	3.3			21	0.5
		61	3.1			20	0.2
		60	2.3				
		59	1.4				
		58	1.3				
		57	1.2				
		56	1.1				
		55	1.1				
		54	1.1				
		53	1.1				
		52	1.1				
		51	1.1				
		50	1.1				
		49	1.1				
		48	1.0				
		47	0.9				
		46	0.9				
		45	0.4				

Table 5.11, continued

Estimated National Norms for MLA Proficiency Test Skill Scores, College Senior FL Majors (Percentile Ranks of Converted Scores)

RUSSIAN

Listening		Speaking		Reading		Writing	
Score	%ile Rank	Score	%ile Rank	Score	%ile Rank	Score	%ile Rank
55	97.7	115	98.8	65	99.4	80	98.8
54	95.5	114	97.6	64	98.7	79	97.6
53	95.5	113	97.6	63	98.7	78	96.5
52	94.4	112	97.6	62	98.7	77	93.7
51	90.9	111	97.6	61	98.3	76	91.6
50	87.3	110	97.6	60	97.5	75	90.1
49	84.2	109	97.6	59	95.4	74	87.0
48	80.0	108	97.6	58	93.2	73	83.5
47	75.6	107	96.3	57	91.3	72	79.7
46	71.7	106	95.1	56	90.0	71	76.1
45	66.4	105	95.1	55	89.6	70	73.8
44	58.1	104	94.4	54	87.9	69	70.6
43	51.5	103	93.8	53	86.6	68	63.6
42	44.3	102	93.8	52	85.9	67	56.5
41	38.9	101	93.2	51	84.8	66	53.8
40	35.6	100	91.5	50	81.8	65	49.1
39	31.2	99	90.4	49	79.3	64	44.0
38	28.4	98	90.4	48	76.0	63	43.0
37	24.1	97	89.7	47	71.6	62	41.9
36	19.2	96	87.8	46	69.0	61	39.1
35	15.0	95	86.1	45	66.9	60	36.4
34	7.9	94	85.3	44	64.6	59	33.7
33	2.3	93	84.9	43	61.3	58	30.1
32	1.2	92	84.9	42	58.6	57	28.2
31	0.4	91	84.9	41	55.2	56	27.5
		90	84.1	40	52.1	55	25.8
		89	82.8	39	49.4	54	23.7
		88	80.4	38	45.3	53	22.2
		87	78.0	37	39.7	52	18.2
		86	74.7	36	33.2	51	13.5
		85	71.3	35	29.4	50	11.2
		84	69.2	34	25.7	49	9.5
		83	67.1	33	22.1	48	9.1
		82	65.8	32	20.2	47	9.1
		81	60.3	31	16.6	46	9.1
		80	54.6	30	12.9	45	8.7
		79	52.3	29	10.4	44	8.3
		78	47.0	28	7.8	43	7.5
		77	40.9	27	6.0	42	6.7
		76	37.1	26	4.3	41	6.1
		75	32.9	25	2.7	40	5.2
		74	29.5	24	1.2	39	4.8
		73	26.6			38	4.1
		72	24.0			37	3.1
		71	23.6			36	2.9
		70	22.0			35	1.4
		69	19.7				
		68	18.9				
		67	18.9				
		66	17.4				
		65	14.6				
		64	12.4				
		63	11.6				
		62	11.1				
		61	10.7				
		60	9.7				
		59	7.9				
		58	6.7				
		57	5.9				
		56	4.8				
		55	4.0				
		54	3.7				
		53	2.5				
		52	1.3				
		51	0.9				
		50	0.2				

Table 5.11, continued

Estimated National Norms for MLA Proficiency Test Skill Scores, College Senior FL Majors (Percentile Ranks of Converted Scores)

SPANISH							
Listening		Speaking		Reading		Writing	
Score	%ile Rank	Score	%ile Rank	Score	%ile Rank	Score	%ile Rank
56	99.4	118	99.9	69	100.0	77	100.0
55	98.0	117	99.8	68	99.9	76	99.8
54	95.6	116	99.8	67	99.8	75	99.6
53	92.0	115	99.8	66	99.3	74	99.3
52	87.1	114	99.4	65	98.7	73	98.7
51	81.9	113	99.0	64	97.8	72	98.0
50	77.4	112	98.9	63	96.6	71	97.0
49	73.1	111	98.8	62	95.1	70	95.3
48	68.1	110	98.7	61	93.9	69	93.7
47	62.4	109	98.5	60	93.0	68	91.8
46	57.3	108	98.2	59	91.2	67	89.3
45	52.6	107	97.9	58	88.8	66	87.5
44	48.0	106	97.4	57	86.2	65	85.7
43	42.9	105	96.9	56	83.2	64	83.4
42	37.1	104	96.0	55	80.5	63	80.4
41	32.0	103	95.2	54	77.7	62	76.8
40	27.7	102	94.2	53	74.3	61	73.4
39	24.1	101	92.9	52	70.9	60	69.9
38	20.9	100	91.8	51	67.3	59	66.4
37	18.0	99	90.4	50	62.8	58	63.1
36	14.6	98	89.0	49	58.6	57	59.4
35	11.3	97	87.3	48	54.9	56	55.5
34	8.7	96	85.4	47	50.7	55	52.2
33	7.0	95	83.6	46	46.3	54	48.5
32	5.7	94	82.2	45	41.9	53	44.6
31	4.1	93	80.5	44	37.8	52	41.7
30	2.5	92	78.1	43	33.9	51	39.0
29	1.5	91	75.0	42	29.5	50	36.4
28	0.9	90	71.9	41	25.2	49	33.9
27	0.4	89	68.1	40	21.2	48	31.1
26	0.2	88	64.0	39	17.6	47	27.7
25	0.1	87	59.7	38	14.1	46	24.1
24	0.1	86	55.5	37	11.5	45	21.1
		85	52.4	36	9.4	44	18.3
		84	49.0	35	6.8	43	15.8
		83	45.3	34	4.6	42	13.6
		82	41.8	33	3.1	41	11.2
		81	39.0	32	2.1	40	9.0
		80	36.9	31	1.4	39	7.8
		79	34.0	30	0.7	38	7.0
		78	30.8	29	0.3	37	5.6
		77	28.5	28	0.1	36	4.2
		76	26.3	27	0.0	35	3.3
		75	23.6			34	2.7
		74	21.0			33	2.1
		73	18.4			32	1.5
		72	16.0			31	1.1
		71	14.6			30	0.9
		70	13.3			29	0.7
		69	11.5			28	0.6
		68	10.1			27	0.4
		67	9.0			26	0.3
		66	7.9			25	0.2
		65	7.1			24	0.1
		64	6.2				
		63	5.6				
		62	5.0				
		61	4.2				
		60	3.3				
		59	2.6				
		58	2.2				
		57	1.8				
		56	1.6				
		55	1.5				
		54	1.1				
		53	0.7				
		52	0.6				
		51	0.6				
		50	0.5				
		49	0.3				
		48	0.2				
		47	0.2				
		46	0.1				

Table 5.12 (a)

Estimated College Senior Major National Norms for Selected Percentile Ranks

FRENCH (Est. N = 5564)

Percentile Ranks	Converted Scores Listening	Speaking	Reading	Writing
99	56	97	65	68
97	55	91	64	64
95	54	89	62	62
90	53	85	60	59
85	52	82	58	57
80	51	80	56	55
75	50	79	55	53
70	49	78	53	52
65	48	77	52	50
60	46	75	51	49
55	45	74	49	47
50	44	73	48	46
45	43	72	47	45
40	41	70	46	43
35	40	69	44	42
30	38	68	43	40
25	37	66	41	39
20	35	65	40	37
15	33	63	38	35
10	32	61	36	33
5	30	57	33	30
3	29	55	32	28
1	27	51	29	25
No. cases actually tested	(1162)	(1123)	(1194)	(1193)
Mean (est. norms)	43.08	72.75	47.86	46.05
Standard Deviation (est. norms)	7.87	9.48	8.83	9.77

F7

Table 5.12 (b)

Estimated College Senior Major National Norms for Selected Percentile Ranks

GERMAN (Est. N = 1848)

Percentile Ranks	Converted Scores Listening	 Speaking	 Reading	 Writing
99	56	112	68	74
97	56	109	67	72
95	55	107	66	71
90	53	102	62	68
85	52	100	60	65
80	51	96	59	63
75	50	95	58	62
70	49	93	56	60
65	48	91	55	58
60	47	89	54	57
55	45	88	52	54
50	44	86	50	51
45	43	85	49	48
40	40	83	48	46
35	38	81	45	44
30	36	79	44	43
25	34	78	43	40
20	33	75	41	37
15	32	72	39	35
10	30	68	36	31
5	28	65	33	27
3	28	61	32	25
1	27	48	30	24
No. cases actually tested	(337)	(326)	(336)	(337)
Mean (est. norms)	42.55	85.75	49.85	50.42
Standard Deviation (est. norms)	8.76	12.91	9.78	13.51

Table 5.12 (c)

Estimated College Senior Major National Norms for Selected Percentile Ranks

RUSSIAN (Est. N = 481)

Percentile Ranks	Converted Scores			
	Listening	Speaking	Reading	Writing
99	55	115	64	80
97	55	111	60	79
95	53	106	59	77
90	51	99	56	75
85	49	93	51	73
80	48	88	49	72
75	47	86	48	71
70	46	84	46	69
65	45	82	44	68
60	44	81	43	67
55	44	80	41	66
50	43	79	39	65
45	42	78	38	64
40	41	77	37	61
35	40	76	36	59
30	39	74	35	58
25	37	72	34	55
20	36	69	32	52
15	35	65	31	51
10	34	60	29	49
5	33	56	26	40
3	33	53	25	37
1	32	51	24	35
No. cases actually tested	(90)	(89)	(90)	(88)
Mean (est. norms)	42.53	79.24	40.91	62.54
Standard Deviation (est. norms)	6.03	13.54	9.69	10.93

Table 5.12 (d)

Estimated College Senior Major National Norms
for Selected Percentile Ranks

SPANISH (Est. N = 4439)

Percentile Ranks	Converted Scores Listening	Speaking	Reading	Writing
99	56	113	65	74
97	55	105	63	71
95	54	103	62	70
90	53	99	59	67
85	52	96	57	65
80	51	93	55	63
75	49	91	53	61
70	48	90	52	60
65	47	88	50	59
60	47	87	49	57
55	46	86	48	56
50	44	84	47	54
45	43	83	46	53
40	42	81	45	51
35	42	79	43	49
30	41	78	42	48
25	39	76	41	46
20	38	74	40	45
15	36	71	38	43
10	35	68	36	40
5	32	62	34	37
3	30	60	33	35
1	28	54	30	31
No. cases actually tested	(859)	(820)	(861)	(865)
Mean (est. norms)	43.97	83.52	47.21	53.80
Standard Deviation (est. norms)	6.79	12.02	8.31	10.16

Table 5.13

Comparison of Estimated College Senior FL Major Norms with Pretest and Posttest NDEA Norms, in Terms of Means ($\overline{X}$) and Standard Deviations (σ)

		Converted Score Statistics							
		Listening		Speaking		Reading		Writing	
	N*	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ
				FRENCH					
CFLT Data	5564	43.08	7.87	72.75	9.48	47.86	8.83	46.05	9.77
NDEA Pretest	4665	38.1	8.8	72.1	19.3	43.3	10.8	43.0	12.9
NDEA Posttest	4721	42.8	8.6	81.0	17.8	45.7	10.5	45.7	12.6
				GERMAN					
CFLT Data	1848	42.55	8.76	85.75	12.91	49.85	9.78	50.42	13.51
NDEA Pretest	1099	38.4	9.1	79.5	17.7	44.4	11.9	45.8	16.5
NDEA Posttest	1250	42.1	9.1	85.3	19.4	48.1	10.9	48.3	14.6
				RUSSIAN					
CFLT Data	481	42.53	6.03	79.24	13.54	40.91	9.69	62.54	10.93
NDEA Pretest	573	38.4	7.0	73.0	17.7	35.5	10.7	48.6	16.5
NDEA Posttest	562	43.3	7.5	87.1	17.0	38.2	10.8	53.4	14.8
				SPANISH					
CFLT Data	4439	43.97	6.79	83.52	12.02	47.21	8.31	53.80	10.16
NDEA Pretest	4389	39.3	8.0	72.1	20.4	42.2	10.2	46.2	13.7
NDEA Posttest	4452	41.8	7.6	78.2	17.5	44.6	9.5	50.2	12.8

*For the CFLT Data, the N given is the estimated number of students in the 1965 population. For the NDEA data, N's are the average numbers taking the tests.

in a foreign language. The NDEA groups, on the other hand, were apparently composed of teachers who were very heterogeneous in background--ranging from the accomplished native speaker of the language to the person who has hardly had any training in the foreign language but was, say, transferred from English or science teaching to teach a foreign language. The differences between the CFLT and the NDEA norms are particularly striking in the Speaking tests. In French and Russian, the NDEA post-test Speaking means are higher than the CFLT means; in all four languages, the NDEA standard deviations are much higher than those in the CFLT data.

Evaluation of CFLT Norms against FSI Absolute Proficiency Ratings

Certain conclusions about the CFLT data in relation to FSI "absolute language proficiency" ratings can be drawn by inspection of the figures (5.1 through 5.4). To provide a more exact basis for comparison, we have prepared Table 5.14, which shows the estimated percentages of the CFLT populations attaining or exceeding each FSI rating.

It is interesting to compare the language groups with respect to the proportions attaining or exceeding FSI level "3"--"Minimum professional proficiency" as defined by the Department of State. This level may be regarded as a minimal qualifying level for a teacher of the foreign language. In French, 19.3% are qualified at this level in Listening, only 4.4% in Speaking, 47.9% in Reading, and 45.7% in Writing. In German, more are qualified in the audio-lingual skills than was the case in French, but about the same proportions in Reading and Writing. Specifically, the proportions qualifying in German in the MLA Skills Tests are 47.6% in Listening, 22.1% in Speaking, 49.8% in Reading, and 44.7% in Writing.

Students of Russian are generally poor in comparison to FSI standards. Only 11.3% attain or exceed the "3" level in Listening, 6.2% in Speaking, 17.0% in Reading, and 12.4% in Writing.

Students of Spanish show up reasonably well in comparison to the FSI standard; 33.6% qualify at the "3" level in Listening, 27.8% in Speaking, 58.1% in Reading, and 47.8% in Writing.

The reader will have noticed, no doubt, that over all languages, students are much less well prepared in Listening and Speaking skills than in Reading and Writing.

Comparison of CFLT Norms against Myers-Melton Standards

Some readers will prefer to evaluate the CFLT Norms with reference to the Myers-Melton (1964) qualification standards. Table 5.15 has been prepared to show the absolute numbers and percentages of the CFLT population falling in each of the Myers-Melton categories, which, it will be recalled, were intended to correspond to the qualification levels (Superior, Good, Minimal, and Unsatisfactory) established by a committee of the Modern Language Association in 1954.

In general, the "Superior" level of the Myers-Melton standards is an even more demanding criterion than the "3" level of the FSI ratings. For example, in French, only 16.48% of the population is estimated to qualify as "Superior" in Listening, whereas 19.3% qualify at the "3" level of the FSI ratings. Less than 1% of the French students qualify as Superior in Speaking, and only 25.20% in Reading and 7.56% in Writing. Similar comparisons may be made for the other language groups.

On the other hand, according to the Myers-Melton standards, relatively few of the CFLT population fall into the Unsatisfactory category. (In Russian, the Myers-Melton ratings do not distinguish between the Minimal and Unsatisfactory ratings.)

Still, whether one uses the FSI ratings or the Myers-Melton ratings as standards, the net showing of the CFLT population is not good. The CFLT groups are particularly deficient in audio-lingual skills. Only about a third to a half are at a satisfactory level even in Reading and Writing skills.

Table 5.14

Estimated Percentages of CFLT Population Attaining or Exceeding Each FSI Rating

FSI "S"	Listening	Speaking	FSI "R"	Reading	Writing
			FRENCH		
S-5	*	0.0	R-5	*	0.0
S-4+	*	0.0	R-4+	0.5	1.7
S-4	*	0.0	R-4	8.0	9.7
S-3+	1.3	1.1	R-3+	25.4	24.2
S-3	19.3	4.4	R-3	47.9	45.7
S-2+	40.4	17.5	R-2+	70.4	68.0
S-2	58.8	47.5	R-2	87.5	86.1
S-1+	74.4	76.2	R-1+	98.1	96.3
S-1	88.2	94.8	R-1	100.0	99.8
			GERMAN		
S-5	*	*	R-5	*	*
S-4+	2.0	0.0	R-4+	5.1	1.7
S-4	16.9	0.4	R-4	13.2	11.0
S-3+	34.0	6.1	R-3+	32.1	28.4
S-3	47.6	22.1	R-3	49.8	44.7
S-2+	58.5	50.0	R-2+	65.2	56.1
S-2	67.0	75.9	R-2	80.9	72.7
S-1+	77.0	89.6	R-1+	90.2	85.2
S-1	89.5	98.6	R-1	98.4	94.4
			RUSSIAN		
S-5	*	*	R-5	*	*
S-4+	*	*	R-4+	*	*
S-4	*	0.0	R-4	0.0	*
S-3+	3.4	2.4	R-3+	6.4	*
S-3	11.3	6.2	R-3	17.0	12.4
S-2+	27.5	14.8	R-2+	39.8	46.0
S-2	55.7	27.3	R-2	71.3	69.9
S-1+	72.0	58.5	R-1+	94.0	89.0
S-1	93.8	80.4	R-1	100.0	93.4
			SPANISH		
S-5	*	0.2	R-5	3.3	*
S-4+	*	1.6	R-4+	9.8	0.0
S-4	1.6	5.2	R-4	22.0	2.6
S-3+	15.5	13.9	R-3+	39.3	19.0
S-3	33.6	27.8	R-3	58.1	47.8
S-2+	52.5	51.7	R-2+	76.8	73.0
S-2	71.8	70.4	R-2	90.8	92.4
S-1+	84.0	85.1	R-1+	98.4	99.0
S-1	93.5	93.0	R-1	100.0	100.0

*The FSI rating equivalent exceeds the maximum score on the test.

Table 5.15

Estimated Numbers and Percentages of CFLT Population Falling in Each of the Myers-Melton Qualification Categories

	Listening		Speaking		Reading		Writing	
	N	%	N	%	N	%	N	%
				FRENCH				
Superior	917	16.48	18	0.33	1402	25.20	420	7.56
Good	2758	49.56	1237	22.23	2716	48.81	2639	47.43
Minimal	1347	24.20	4002	71.92	1376	24.73	2359	42.40
Unsatisfactory	543	9.76	29	0.52	70	1.26	145	2.61
Total	5565	100.00	5565	100.00	5565	100.00	5565	100.00
				GERMAN				
Superior	692	37.45	70	3.78	458	24.78	364	19.72
Good	651	35.21	792	42.84	1001	54.17	826	44.70
Minimal	391	21.16	784	42.41	266	14.38	524	28.36
Unsatisfactory	114	6.18	203	10.97	123	6.67	133	7.22
Total	1848	100.00	1848	100.00	1848	100.00	1848	100.00
				RUSSIAN				
Superior	129	26.76	73	15.14	225	46.83	131	27.20
Good	269	55.94	149	31.10	187	38.89	310	64.52
Minimal Unsatisfactory	83	17.30	258	53.76	69	14.28	40	8.28
Total	481	100.00	481	100.00	481	100.00	481	100.00
				SPANISH				
Superior	1532	34.52	1039	23.40	1213	27.33	590	13.29
Good	1913	43.10	2413	54.35	2363	53.23	2539	57.20
Minimal	851	19.18	897	20.21	789	17.77	1204	27.12
Unsatisfactory	142	3.20	91	2.04	74	1.67	106	2.39
Total	4439	100.00	4439	100.00	4439	100.00	4439	100.00

Chapter VI

STUDENT FACTORS IN FOREIGN LANGUAGE SKILL ATTAINMENT

1. Introduction

"Student factors" are those characteristics of students that they bring with them to the instructional program; some of these characteristics may be inherent in them, like sex and age, while others, like motivations and interests, may be modified by their experiences in the course of foreign language study. Since this investigation was a descriptive survey in which all measurements and observations were taken at roughly a single point of time in the life histories of our subjects, it is impossible to draw well-founded inferences as to any causal relationships that may exist between the characteristics of students and their foreign language attainments at the time of graduating from college. We cannot even indicate what changes may have occurred in student characteristics in the course of time. All we can do is to compare different groups of students with respect to foreign language attainments and observe any statistical associations that may exist. Nevertheless, it is of interest to describe different groups, note any differences in foreign language attainments, and attempt to suggest possible reasons for these differences. In any case, the findings are likely to throw light on the interpretation of test scores. For example, if there were marked sex differences in foreign language attainments (actually, there are not) a given score might have a different interpretation depending on the sex of the individual. The findings would also be useful in analyzing the variations that may exist in the sample of graduating foreign language majors studied here.

All analyses reported in this chapter are based on the actual sample of foreign language majors that were tested, or upon sub-samples thereof. No effort is made to estimate the characteristics of students who were not tested, or of the total population of students tested and not tested). The limitations of the sample pointed out in Chapters III and V must therefore be borne in mind in interpreting the findings.

For the most part, fairly standard statistical techniques have been employed to make the analyses. The reader who wishes to learn more about these techniques should consult standard textbooks of statistics, such as McNemar (1962) or Hays (1963). Many of the analyses in this and subsequent chapters were performed on high-speed computers and employed programs reported by Cooley and Lohnes (1962) and Jones (1964).

2. Sex and Age

The analyses to be presented here are based on all seniors complete on the four MLA skills tests and on pages 1 and 4 of the Student Questionnaire giving background information, and who are "regular cases" who had a normal course of training in foreign languages starting either in elementary school, high school, or college. Data for Italian were generally not analyzed because of the small numbers of cases. As far as can be determined, there is little bias in the sample with respect to sex (See Chapter III). As was pointed out earlier, the majority of foreign language majors are females; in the particular sample used here, males constitute only 15.9% of the French majors, 41.0% of the German majors, 37.5% of the Russian majors, and 24.8% of the Spanish majors.

Males' and females' score distributions on each skill test in each language were compared by means of the *t*-test, with results shown in Table 6.1. Only one difference

Table 6.1

MLA Skills Scores by Sex

Skill	Males			Females			Total			
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	t
					FRENCH					
Listening	177	43.8	8.4	937	44.2	7.5	1114	44.1	7.6	-0.68
Speaking	"	74.6	10.2	"	73.4	9.0	"	73.6	9.2	1.62
Reading	"	49.2	10.0	"	48.7	8.3	"	48.8	8.6	0.60
Writing	"	46.2	10.7	"	47.3	9.3	"	47.2	9.5	-1.49
					GERMAN					
Listening	133	43.3	8.5	191	43.5	8.7	324	43.4	8.6	-0.17
Speaking	"	87.4	12.2	"	86.3	13.2	"	86.8	12.8	0.80
Reading	"	51.4	10.1	"	50.8	9.5	"	51.0	9.8	0.55
Writing	"	52.4	13.1	"	51.8	13.5	"	52.1	13.3	0.39
					RUSSIAN					
Listening	33	41.7	6.0	55	43.1	5.9	88	42.6	6.0	-1.07
Speaking	"	78.9	14.2	"	78.0	11.6	"	78.3	12.7	0.32
Reading	"	40.1	8.7	"	42.0	10.1	"	41.3	9.6	-0.85
Writing	"	61.9	11.3	"	62.2	10.7	"	62.1	10.9	-0.11
					SPANISH					
Listening	202	44.8	10.3	612	44.5	6.7	814	44.5	6.8	0.37
Speaking	"	86.0	12.3	"	83.4	11.8	"	84.0	12.0	2.74**
Reading	"	48.0	8.6	"	47.7	8.3	"	47.8	8.4	0.41
Writing	"	54.2	10.8	"	54.8	9.8	"	54.6	10.1	-0.72

**Significance ($p < .01$)

was significant beyond the 5% level: in the Spanish group, males were significantly superior in Speaking, at the 1% level of significance (two-tailed test). We have no explanation for this and since all the other differences were only what could be expected by chance, it may be regarded as a "chance" significant difference. There was not even a pattern over languages to suggest that one sex tended to be superior in some skills but not in others. As far as our sample is concerned, males and females do not differ in average foreign language skill attainment.

The students were asked to state their "age." (No specification such as "age at nearest birthday" or "age at last birthday" was asked for in the questionnaire.) Table 6.2 shows the distribution of ages stated, by language. The modal age in all languages is 21; 88.8% of the ages are comprised within 20, 21, and 22. The distributions are nevertheless highly skewed; while there is only a scattering of ages 18 and 19, there are substantial numbers of students graduating in the late 20's, the 30's, and 40's. In each language, the group was broken down into three subgroups, 20 or below, 21 and 22, and 23 and over, and these subgroups were compared by one-way analysis of variance with respect to their mean MLA skills scores, with results as shown in Table 6.3. Many of these comparisons approached significance at the 1% level, but there are no clear patterns. In French and Spanish there is a slight tendency for the younger graduates to have superior scores. This may possibly foreshadow the finding (Chapter VII) that those who started foreign language study early had higher scores than those who started later. In any case, the age groups are in general so close together in FL attainment that no recommendations concerning this variable can be given. There is certainly no strong tendency for older students to obtain inferior scores.

There is a slight tendency (the data are not shown in the tables) for disproportionately more of the older students (23 and over) to be males.

3. Foreign Language Aptitude

Students tested in the national study are on the average quite high on both the part and the total scores of the Modern Language Aptitude Test (MLAT). Table 6.4 shows the means and standard deviations of these scores for four language groups, as well as the limited normative data available from the Manual (1959) of the MLAT. There are three possible factors making for superior average scores for the CFLT samples:

(1) Norms shown in the manual for high school (Grades 9, 10, 11 and college freshman groups exhibit monotonically increasing mean scores. For example, "short form" (total, parts 3-5) means for females for fairly large norm groups are as follows:

	N	Mean	S.D.
Grade 9	391	43.3	13.2
Grade 10	304	46.8	15.6
Grade 11	233	53.4	14.5
College Freshmen	277	61.4	17.3

On the assumption that growth continues throughout the college years, it is not unreasonable to expect average scores for seniors to be somewhat higher than the College Freshmen norms. As may be seen in Table 6.4, the distribution of MLAT Total scores for all 2364 cases in the CFLT sample has a mean of 76.2 and a standard deviation of 17.4.

(2) The norms shown in the Manual and cited above are for unselected students taking foreign language courses. It could be expected that students majoring in foreign

Table 6.2

Age Distributions of FL Majors at Graduation, by Language

	LANGUAGE											
	FRENCH		GERMAN		ITALIAN		RUSSIAN		SPANISH		TOTAL	
Age	f	cum. %	f	cum. %	f	cum. %	f	cum. %	f	cum. %	f	cum. %
18	0	0.0	0	0.0	0	0.0	0	0.0	1	0.1	1	0.04
19	7	0.6	0	0.0	0	0.0	0	0.0	5	0.7	12	0.5
20	81	7.9	15	4.6	1	4.2	8	9.1	51	7.0	156	7.1
21	686	69.5	175	58.6	18	79.2	45	60.2	448	62.0	1372	65.2
22	256	92.5	88	85.8	3	91.7	25	88.6	201	86.7	573	89.4
23	30	95.2	14	90.1	0	91.7	2	90.9	37	91.3	83	92.9
24	15	96.5	11	93.5	0	91.7	3	94.3	19	93.6	48	95.0
25	11	97.5	8	96.0	0	91.7	0	94.3	11	95.0	30	96.2
26	3	97.7	5	97.5	0	91.7	3	97.7	8	95.9	19	97.0
27	2	97.9	3	98.5	1	95.8	0	97.7	8	96.9	14	97.6
28	2	98.1	2	99.1	0	95.8	0	97.7	5	97.5	9	98.0
29	2	98.3	1	99.3	0	95.8	0	97.7	3	97.9	6	98.3
30-34	2	98.5	1	99.7	1	100.0	0	97.7	8	98.9	12	98.8
35-39	6	99.0	0	99.7	0		0	97.7	2	99.1	8	99.1
40-44	8	99.7	1	100.0	0		1	98.9	4	99.6	14	99.7
45-49	2	99.9	0		0		0	98.9	1	99.8	3	99.8
50+	1	100.0	0		0		1	100.0	2	100.0	4	100.0
Total	1114		324		24		88		814		2364	

Table 6.3

Mean MLA Skill Test Scores by Age Groups

Language	Age Group	N	Listening	Speaking	Reading	Writing
FRENCH	20 or less	88	46.4	76.2	50.2	50.0
	21-22	942	43.9	73.2	48.6	46.9
	23+	84	43.7	74.3	49.6	47.0
	Total	1114				
	P		<.02	<.02	.18	<.02
GERMAN	20 or less	15	42.9	88.3	52.0	55.1
	21-22	263	43.0	85.8	50.6	51.5
	23+	46	46.3	91.7	53.0	54.6
	Total	324				
	P		.06	<.02	>.10	>.10
RUSSIAN	22 or less	78	42.2	77.5	41.0	61.6
	23+	10	45.1	84.9	43.1	65.9
	Total	88				
	P		>.10	.08	>.10	>.10
SPANISH	20 or less	57	46.5	82.2	50.0	58.1
	21-22	649	44.3	83.7	47.4	54.2
	23+	108	44.6	86.7	49.0	55.0
	Total	814				
	P		.06	<.03	<.02	<.02

Table 6.4

Modern Language Aptitude Test scores, by language, as compared with norms

Samples: All tested students complete on MLA Skill
Scores and Parts 3, 4, & 5 of MLAT,
excluding non-seniors and Italian majors

			National CFLT data MLAT				Norms from Manual*		
Language	N		Pt. 3	Pt. 4	Pt. 5	Total 3-5	N		
FRENCH	1084	M:	27.4	32.1	19.6	79.1	55	M:	61.7
		SD:	9.9	6.5	5.0	15.9	(Women)	SD:	16.1
GERMAN	349	M:	24.1	31.3	18.8	74.2	32	M:	54.6
		SD:	9.9	7.6	5.8	18.1	(Women)	SD:	19.5
RUSSIAN	94	M:	28.5	32.0	19.6	80.1	73	M:	68.9
		SD:	12.0	7.9	5.6	20.4	(Men)	SD:	19.1
SPANISH	837	M:	24.7	29.7	18.6	73.0	39	M:	57.8
		SD:	10.0	7.8	5.7	17.8	(Women)	SD:	16.2
TOTAL	2364	M:	26.0	31.1	19.1	76.2	———		
		SD:	10.1	7.3	5.4	17.4	———		

*These norms are from data shown for "College Sophomores, Juniors, and Seniors" on the "Short Form" of the MLAT, from the Manual, p. 11. The "Short Form" is the same as Parts 3-5 and therefore these data are to be compared with the total scores on Parts 3-5 for the CFLT samples.

languages might have higher average scores than samples of students who are not necessarily "majoring" in a foreign language.

(3) It is possible that foreign language study as such teaches certain skills and knowledges tested by the MLAT and therefore that persons with large amounts of foreign language study will attain higher scores than persons with lesser amounts of such study.

The present study offers no direct evidence as to the magnitude of any of these factors. Chapter VII will offer some indirect evidence that suggests that the third of these factors has little effect; it will be shown, that is, that persons with different amounts of foreign language study do not differ much, on the average, in MLAT scores.

Another interesting aspect of Table 6.4 is the comparison of average MLAT scores across languages. The mean total scores rank as follows:

Russian	80.1
French	79.1
German	74.2
Spanish	73.0

An analysis of variance of these data yields an F-ratio that is highly significant. The results tend to agree with results obtained elsewhere in research on the MLAT: those who select a "difficult" language like Russian tend to have higher foreign language aptitude, and those who select an allegedly "easy" language like Spanish tend to have lower aptitude scores. Apparently there is some degree of self-selection of languages according to language aptitude; it is known from previous research (Carroll, 1932) that individuals' ratings of their own language aptitude tend to correlate with their scores on the MLAT and therefore it is not unreasonable to think that students can and do select languages to study in accordance with their own images of their language aptitude.

Although the average MLAT scores of the CFLT group are high, the score distributions show considerable range, as indicated by the standard deviations, which are comparable with those of the distributions of norming groups discussed in the MLAT manual. The question can be raised, now, whether language aptitude as measured by the MLAT shows any significant relationship with foreign language attainment. A subsidiary but related question is whether the several parts of the MLAT--or at least the three parts used in this study--measure different aspects of language aptitude. Relevant data are shown in Table 6.5.

The intercorrelations of parts of the MLAT are relatively low; this finding accords with results of previous research (Carroll and Sapon, 1958). The parts do indeed measure different aspects of language aptitude.

The correlations of MLAT scores with MLA skill test scores are generally very low; some of them are negative. Only in the case of the French and Spanish groups are they generally positive, but even then, they barely reach statistical significance. These findings, taken at face value, would seem to show that language aptitude has very little to do with success in learning a language. It would be hasty to draw such a conclusion, however, for it must be remembered that the groups studied here are very heterogeneous in their backgrounds. From previous research comes the clear suggestion that language aptitude is related to the amount of time the individual will have to take to acquire a specified degree of competence. These groups are very heterogeneous in the amount of time they have studied their foreign languages; as will be shown in Chapter VII, the cases vary dramatically with respect to how early in their student careers they started studying the foreign languages in which they are majoring, and also with respect to the amount of opportunity they have had to acquire competence through foreign travel and study and through exposure to the foreign language at home. It is only after all these factors are controlled, by statistical means, that it will be possible to examine the

Table 6.5

Intercorrelations of MLAT Part Scores and Correlations with MLA Skill Scores

(Same samples as in Table 6.4)

FRENCH (N = 1084)

				MLA Skill Score			
	Part 3	Part 4	Part 5	Listening	Speaking	Reading	Writing
MLAT Part 3 (Spelling Clues)	1.00	.30	.24	.20	.16	.25	.23
MLAT Part 4 (Words in Sentences)	.30	1.00	.35	.11	.05	.22	.25
MLAT Part 5 (Paired Associates)	.24	.35	1.00	.10	.01	.16	.14
MLAT Total, Parts 3-5	.82	.71	.61	.21	.13	.30	.29

GERMAN (N = 349)

	Part 3	Part 4	Part 5	Listening	Speaking	Reading	Writing
MLAT Part 3	1.00	.43	.32	-.08	-.13	-.06	-.07
MLAT Part 4	.43	1.00	.44	-.12	-.23	-.04	.02
MLAT Part 5	.32	.44	1.00	-.14	-.26	-.16	-.16
MLAT Total, Parts 3-5	.83	.79	.68	-.14	-.25	-.10	-.08

RUSSIAN (N = 94)

	Part 3	Part 4	Part 5	Listening	Speaking	Reading	Writing
MLAT Part 3	1.00	.42	.48	-.15	-.26	-.19	-.20
MLAT Part 4	.42	1.00	.37	.06	-.01	.07	.06
MLAT Part 5	.48	.37	1.00	-.13	-.20	-.16	-.14
MLAT Total, Parts 3-5	.88	.74	.70	-.10	-.21	-.12	-.13

SPANISH (N = 837)

	Part 3	Part 4	Part 5	Listening	Speaking	Reading	Writing
MLAT Part 3	1.00	.38	.26	.14	.07	.15	.20
MLAT Part 4	.38	1.00	.38	.00	-.08	.02	.07
MLAT Part 5	.26	.38	1.00	.00	-.06	.03	.04
MLAT Total, Parts 3-5	.81	.77	.63	.08	-.02	.10	.16

true relationships between language aptitude and the attainment of foreign language skill. This is attempted in Chapter VIII, and it will be shown there that language aptitude has significant positive relationships to skill attainment when the effects of extraneous variables are eliminated.

4. Plans, Interests, and Motives in Foreign Language Study

In the student questionnaire, the following question appeared:

1) For what reason(s) did you decide to concentrate in the language in which you are being tested? Check as many as apply:

- ☐ I expect to teach a foreign language or languages. If so, check the level or levels at which you plan to teach:
    - ☐ at the elementary school level
    - ☐ at the secondary school level
    - ☐ a college, university, or graduate school
    - ☐ other (specify)________________________
- ☐ I expect to use my foreign language knowledge in my future employment or profession. If so, specify the kinds of work you are contemplating:
    - ☐ Linguistics
    - ☐ Business activity (importing, foreign trade, etc.)
    - ☐ Government work (foreign service, United Nations, A.I.D., etc.)
    - ☐ Military service
    - ☐ Work as a translator or interpreter
    - ☐ Other professional work (medicine, science, library work, etc.)
- ☐ I expect to use my foreign language knowledge in travel or study abroad.
- ☐ I majored in a foreign language mainly for interest and enjoyment.
- ☐ I majored in a foreign language because I am particularly interested in the literary and cultural aspects of the language.
- ☐ I am interested in intercultural understanding and communication.
- ☐ No particular reason other than the fact that I needed a major subject.
- ☐ Other reasons (specify) ______________________________________.

The responses of 2340 French, German, Russian, and Spanish majors whose cases were complete on MLA skill test scores are tabulated in Table 6.6. The results for the major categories depict the main reasons why the students thought they were majoring in a foreign language.

Because of the structure of the question, which permitted various combinations of multiple responses, it is not possible to set up exclusive categories for students' reasons for majoring in a foreign language. About half of those planning to use the foreign language in future employment also indicated that they planned to teach.

Overall, 64.6% of the majors plan to go into the teaching of foreign languages, at some level. A slightly higher proportion of females (65.6%) are interested in teaching than are males (61.1%) The four languages are more or less equal in this respect, Spanish being the language where a higher proportion (67.6% are attracted to teaching, and Russian being the language where the smallest proportion (51.1%) are attracted to teaching. Russian is also different from the other languages in that more males than females plan to teach it.

Table 6.7 shows a detailed analysis of the groups planning teaching. Teaching at the secondary school level draws by far the largest percentage of responses -- 85.3%; females (89.3%) more than males (71.2%). The languages rank as follows: Spanish (89.4%), French (86.0%), German (78.0%), Russian (55.6%). Relatively few males (27.8%) plan

Table 6.6

Chief Reasons Given for Majoring in a Foreign Language, By Sex, in Percentages*

Sample: 2340 French, German, Russian, and Spanish Majors
Complete on MLA Skills Test and Questionnaire Responses

Plans for Use of FL, or Reason for Majoring	Language												Total for Four Language Groups		
	French			German			Russian			Spanish					
	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Base N	177	937	1114	133	191	324	33	55	88	202	612	814	545	1795	2340
Teaching	61.0	64.9	64.3	59.4	63.4	61.7	54.5	49.1	51.1	63.4	69.0	67.6	61.1	65.6	64.6
Future Employment	40.1	35.0	35.8	39.1	36.1	37.3	57.6	45.4	50.0	40.6	35.5	36.7	41.1	35.6	36.9
Travel or Study Abroad	68.9	73.9	73.6	73.7	71.7	72.5	72.7	47.3	56.8	62.9	73.9	71.1	68.1	72.9	71.8
Mainly for Interest and Enjoyment	48.6	65.8	63.1	54.9	65.4	61.1	42.4	65.4	56.8	46.0	61.9	58.0	48.8	64.5	60.8
Interest in Lit. and Cultural Aspects	46.3	53.4	52.3	49.6	55.5	53.1	39.4	50.9	46.6	37.1	48.2	45.5	43.3	51.8	49.8
Interested in Inter-cultural Understanding	51.4	59.8	58.4	48.1	53.4	51.2	60.6	56.3	57.9	52.5	63.7	60.9	51.6	60.3	58.3
Needed a Major	2.8	1.9	2.1	6.8	2.1	4.0	--	--	--	4.5	1.6	2.3	4.2	1.8	2.4

*Percentages may add to more than 100% because of overlapping categories and multiple responses.

3

Table 6.7

Levels at Which Majors Plan to Teach, By Sex, In Terms of Percentage*

Sample: As for Table 6.6

	Language												Total for Four Language Groups		
	French			German			Russian			Spanish					
	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Base N (All Majors)	177	937	1114	133	191	324	33	55	88	202	612	814	545	1795	2340
No. of Majors Planning to Teach (Base of %'s below)	108	608	716	79	121	200	18	27	45	128	422	550	333	1178	1811
Level:															
Elementary School	5.6	17.3	15.5	1.3	9.1	6.0	--	--	--	2.3	11.8	9.6	3.0	14.1	11.6
Secondary School	67.6	89.3	86.0	68.4	84.3	78.0	27.8	74.1	55.6	82.0	91.7	89.4	71.2	89.3	85.3
College, etc.	57.4	23.5	28.6	65.8	38.0	49.0	83.3	59.3	68.9	50.0	21.3	28.0	57.9	25.0	32.3
Other	0.9	3.0	2.6	--	--	--	--	--	--	0.8	3.8	3.1	0.6	2.9	2.4
Per cent planning to qualify for a teaching certificate	67.6	98.0	93.4	77.2	92.5	86.5	44.4	70.4	60.0	91.4	99.8	97.8	77.8	97.5	93.1

*Percentages may sum to more than 100% because of multiple responses.

teaching Russian at this level, however. Teaching at the college, university, or graduate school level attracts 32.3% of those planning to teach; it attracts males far more strongly (57.9%) than females (25.0%), a fact that undoubtedly reflects employment opportunities. Teaching at the elementary school level is included in the plans of only 11.6% of the cases, and most of these are females planning to teach either French or Spanish.

Table 6.7 also shows the proportions, out of those planning to teach, who indicated in Question 3 on Page 4 of the Student Questionnaire that they planned to "qualify for a license or certificate for teaching in public schools." These percentages closely parallel, or are a little larger than, the percentages planning to teach at the secondary school level. Evidently most of those planning to teach at the secondary school level also realize the necessity or advantage of having a teaching certificate, and there are also others who feel that a teaching certificate would be a good thing to have "in case."

Table 6.8 gives an analysis of specific employment plans for those thinking of using foreign language skills in work other than teaching. More than half of the students plan on government work (particularly those in Russian), and almost as many plan on work in business (except those in Russian). A surprising proportion (39.0%) think of work in translating or interpreting, particularly females studying Russian. Linguistics, and even less, military service, concerns small numbers. Other professional work such as in medicine, science, library work, etc. attracts small but appreciable proportions.

Returning to Table 6.6, we see that large proportions of all groups are interested in the "fringe benefits" of foreign language study--in particular, use in travel or study abroad (71.8%). 60.8% of the cases claim that they majored in a foreign language because of "interest and enjoyment"--a finding which will no doubt please the foreign language teacher. Roughly half (49.8%) express interest in the literary and cultural aspects of foreign language study, and 58.3% claim interest in the implications of foreign language study for intercultural understanding. Very small proportions take foreign language merely because they "needed a major," from which we may conclude that nearly all the students had serious reasons for foreign language study.

Another aspect of interest in foreign language study is the relative degrees of importance attached to the four basic language skills, listening, speaking, reading, and writing. Students' opinions in this area are related in considerable degree to their vocational plans and to other reasons for majoring in a foreign language. To make analyses of these opinions, male and female students were pooled over foreign language major groups (except for Italian) and the mean "importance ratings" attached to the four skills were computed for subgroups indicating specific vocational goals or other reasons for foreign language study. The results are shown in Table 6.9.

Over all students (see data for "Total" at end of the section entitled "By Intentions to Teach") all four skills were generally rated as "of great importance." On the average Speaking was rated as of most importance, and Listening was a close second. Reading had the third highest importance rating, and Writing had the lowest average importance rating, even though the figure (2.66) was still in the range of "great importance" on the scale. All four skills were rated higher by those planning to teach than by those not planning to teach.

Those planning to teach only at the elementary school level tended to give somewhat lower importance ratings to all four skills than those planning to teach at other levels. They were particularly inclined to give lower importance ratings to Reading and Writing. This finding possibly reflects the fact (as will be shown below) that those planning to teach at the elementary level tend to have somewhat lower MLA skill test scores than those planning to teach at higher educational levels. Furthermore, importance ratings tend to be somewhat correlated with MLA skill scores; on this point, see below.

Ratings of the importance of Reading and Writing tend to be particularly high for those planning to teach at the college level, and these are the students who tend to have the highest average MLA skills scores, as shown in Table 6.12.

Table 6.8

Specific Employment Prospects of Those Planning
To Use FL Skills in Work other than Teaching,
By Sex, In Terms of Percentages*

	Language												Total for Four Language Groups		
	French			German			Russian			Spanish					
	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Base N (All Majors)	177	937	1114	133	191	324	33	55	88	202	612	814	545	1795	2340
No. of Majors Planning Employment other than Teaching (Base of %'s Below)	71	328	399	52	69	121	19	25	44	82	217	299	224	639	863
Type of Work:															
Government work (foreign service, etc.)	50.7	61.0	59.1	42.3	56.5	50.4	73.7	88.0	81.8	46.3	57.6	54.5	49.1	60.4	57.5
Business	29.6	49.3	45.9	36.5	26.0	30.6	--	--	--	40.2	49.8	47.1	32.6	45.1	41.8
Translating and Interpreting	32.3	43.3	39.1	26.9	40.6	34.7	21.0	80.0	54.6	21.9	44.7	38.4	26.3	43.5	39.0
Linguistics	21.1	14.9	16.0	23.1	14.4	18.2	21.1	24.0	22.7	19.5	12.4	14.4	21.0	14.4	16.1
Military Service	21.1	1.2	4.8	26.9	1.4	12.4	--	--	--	18.3	0.9	5.7	19.6	1.1	5.9
Other prof. work	22.5	14.6	16.0	17.3	23.2	20.7	5.3	36.0	22.7	28.0	21.2	23.1	21.9	18.6	19.5

*Percentages may sum to more than 100% because of multiple responses.

Table 6.9

Mean Importance Ratings Given to Foreign Language Skills by Students

Sample: 2340 FL Majors ("Regular Cases") in All Languages Except Italian

Coding of Scale: 3 = great importance
2 = some importance
1 = little importance

Group	N	%	Listen.	Speak.	Read.	Write
By Intentions to Teach						
Planning to Teach:						
-At Elementary Level Only	59	2.52	2.81	2.93	2.58	2.42
-At Elem. or Sec. Levels	99	4.23	2.95	2.97	2.86	2.74
-At Secondary Level Only	964	41.20	2.90	2.97	2.80	2.71
-At Sec. or College Level	196	8.38	2.92	2.95	2.90	2.80
-At College Level Only	270	11.54	2.91	2.95	2.96	2.82
-Other level, or no level given	64	2.74	2.91	2.95	2.82	2.72
All Planning to Teach	1652	70.60	2.90	2.96	2.83	2.73
All Not Planning to Teach	688	29.40	2.82	2.77	2.74	2.49
Total	2340	100.00	2.88	2.91	2.81	2.66
By vocational goals						
All Planning to Use FL in Future Employment (other than teaching)	863	36.88	2.91	2.89	2.79	2.65
-In Government Work	496	21.20	2.94	2.92	2.79	2.66
-In Business	368	15.73	2.92	2.90	2.75	2.65
-In Translating or Interpreting	337	14.40	2.94	2.94	2.84	2.75
-In Linguistics	139	5.94	2.91	2.92	2.85	2.75
-In Military Service	55	2.35	2.95	2.91	2.69	2.55
-In Other Professional Work	168	7.18	2.81	2.76	2.79	2.51
By Other Reasons for FL Study						
Use FL in travel	1679	71.75	2.92	2.94	2.82	2.69
"Interest and Enjoyment"	1423	60.81	2.89	2.90	2.82	2.67
Interest in foreign culture	1166	49.83	2.90	2.90	2.88	2.72
Interested in intercultural understanding	1364	58.29	2.93	2.93	2.83	2.70
"Needed a Major"	58	2.48	2.67	2.72	2.60	2.33

Students planning to use their foreign language skills in some future employment other than teaching (whether or not they in addition plan to teach) tend to give slightly higher importance ratings to Listening, and slightly lower ratings to the other three skills, as compared to the total group. However, it is among those planning to use foreign language knowledge in translating or interpreting, or in linguistics, that Reading and Writing attract the highest importance ratings.

All these findings give strong support to contemporary policies in foreign language teaching whereby listening and speaking are accorded emphasis and attention at least equal to that accorded to reading and writing. Most students seem to regard reading and writing as "advanced" skills of great importance, particularly for specialized work such as college teaching, translating and interpreting, and linguistics, but listening and speaking are regarded as being of even greater importance, particularly for teaching at elementary and secondary school levels.

Nevertheless, students do differ in the relative importance they attach to listening and speaking, on the one hand, and reading and writing, on the other. In Table 6.10, where (among other things) the intercorrelations of importance ratings are given for four language groups, we see that importance ratings of listening and speaking, and also the importance ratings of reading and writing, are much more highly correlated with each other than with the other pair of skills. Further, the language groups differ somewhat in their mean ratings. French, German, and Spanish groups characteristically tend to attach more importance to listening and speaking than to reading and writing (the contrast is particularly strong in Spanish), while the Russian group attaches most importance to reading, less to listening and speaking, and least to writing. These differences reflect well-known facts about the use of these languages: American students in French and German are interested about equally in oral communication with speakers of those languages and in their literatures; students of Spanish are interested more in oral communication than in Spanish literature, and students of Russian are more interested in reading literature (either belles-lettres, or scientific writings, or historical and political literature) than they are in oral communication with speakers of Russian.

Let us re-emphasize at this point, however, that even within language groups students differ in the relative importance they attach to oral and to written skills. To what extent are these differences related to their attainments in these skills as measured by the MLA skills tests? Table 6.10 contains also the correlations of importance ratings with MLA skills scores. First we note that all these correlations are positive, indicating that the general degree of importance a student attaches to any foreign language skill is somewhat related to his foreign language attainments in general. We might expect also, however, that the student who attaches greater importance to oral skills will have made greater attainment in these skills, and likewise, the student who attaches greater importance to written language skills will do better, relatively, in reading and writing tests. Table 6.11 presents a special analysis that sought to investigate this matter. It utilizes a rather rarely employed statistical technique known as canonical analysis. (For a fairly detailed explanation of this technique, see Chapter VIII, Section 4.) The results are clearest in French, where there was the largest number of cases. The first canonical correlation for that group, .24, was highly significant and provides a measure of the fact that (as indicated by the canonical regression weights) there is generally a correlation between importance ratings and skills test scores. That is, if we were to add up the importance ratings, with all positive weights, we could predict a linear combination of test scores--a linear combination that would weight the speaking and writing skill scores. This finding merely confirms, with precision, the observation made above that the importance ratings and skills scores are all positively correlated, and it tends to confirm our interpretation of this observation.

It is the second canonical correlation that is of most interest, for this has to do with relationships between importance ratings and skill scores after the variance due to the first canonical correlation has been "extracted," that is, statistically controlled or taken account of. In the French group, this second canonical correlation is .12--a rather small value, but nevertheless statistically significant beyond the 1% level of confidence. Looking at the canonical regression weights, we notice that there is a

Table 6.10

Intercorrelations of FL Skill Importance Ratings and Correlations with MLA Skill Scores, By Language

N = 2340 "Regular" Cases in Four Languages

		Importance Ratings				MLA Skill Scores			
		L	S	R	W	L	S	R	W
					FRENCH (N = 1114)				
Mean		2.89	2.91	2.82	2.67	44.11	73.55	48.81	47.15
S.D.		.32	.30	.42	.54	7.62	9.25	8.59	9.45
r's	L	1.00	.51	.14	.25	.14	.18	.10	.14
	S	.51	1.00	.05	.31	.08	.16	.05	.08
	R	.14	.05	1.00	.54	.14	.14	.16	.14
	W	.25	.31	.54	1.00	.13	.17	.16	.18
					GERMAN (N = 324)				
Mean		2.83	2.82	2.81	2.56	43.44	86.75	51.03	52.09
S.D.		.45	.49	.42	.61	8.62	12.80	9.75	13.32
r's	L	1.00	.78	.29	.43	.23	.25	.17	.24
	S	.78	1.00	.26	.51	.19	.21	.13	.19
	R	.29	.26	1.00	.48	.26	.20	.26	.27
	W	.43	.51	.48	1.00	.21	.23	.19	.23
					RUSSIAN (N = 88)				
Mean		2.75	2.77	2.89	2.52	42.56	78.34	41.28	62.10
S.D.		.48	.45	.35	.56	5.99	12.67	9.64	10.89
r's	L	1.00	.53	-.10	.36	.14	.01	.16	.07
	S	.53	1.00	-.02	.47	.11	.05	.08	.09
	R	-.10	-.02	1.00	.24	.18	.21	.11	.35
	W	.36	.47	.47	1.00	.05	.05	.05	.14
					SPANISH (N = 814)				
Mean		2.90	2.94	2.77	2.70	44.47	84.01	47.82	54.60
S.D.		.33	.27	.45	.51	6.79	12.00	8.37	10.11
r's	L	1.00	.46	.18	.13	.11	.08	.08	.11
	S	.46	1.00	.16	.19	.04	.02	.00	.05
	R	.18	.16	1.00	.59	.10	.10	.13	.15
	W	.13	.19	.59	1.00	.14	.13	.12	.16

Table 6.11

First Two Canonical Correlations between FL Skill Importance Ratings and MLA Skill Scores, and Canonical Regression Weights

	Ist Canonical Correlation		IInd Canonical Correlation	
	Importance Ratings	MLA Skill Scores	Importance Ratings	MLA Skill Scores
	FRENCH (N = 1114)			
	R_I = .24**		R_{II} = .12**	
Listening	.62	-.12	.05	.17
Speaking	.31	.90	.81	.63
Reading	.44	.09	-.14	-.70
Writing	.57	.41	-.56	-.28
	GERMAN (N = 324)			
	R_I = .34**		R_{II} = .18	
Listening	.71	.43	.22	-.01
Speaking	-.04	.32	.39	.66
Reading	.61	-.37	-.84	-.75
Writing	.34	.76	.29	.03
	RUSSIAN (N = 88)			
	R_I = .41		R_{II} = .22	
Listening	-.11	-.12	.86	.75
Speaking	.16	.01	.13	-.37
Reading	.97	-.39	.22	.46
Writing	.15	.91	-.44	-.30
	SPANISH (N = 814)			
	R_I = .20**		R_{II} = .11	
Listening	.62	.37	.21	.77
Speaking	-.15	.19	.17	.00
Reading	.27	-.25	-.83	-.60
Writing	.73	.87	.49	-.22

**p < .01

pairing of signs such that in making the best prediction of skill scores from importance ratings after the first canonical variance has been taken account of, we would weight both importance ratings and skills scores in Listening and Speaking positively, and both importance ratings and skills scores in Reading and Writing negatively. In other words, the difference between importance ratings for oral and written skills predicts a corresponding difference between test scores. This is clear evidence that at least in the French group, students tend to get higher scores in skills to which they attach greater importance. The causal relationship, however, is not specified: it may be that student attitudes cause greater devotion to the attainment of specific skills, or it may be that greater success in particular skills--for whatever reasons--causes the student to attach greater importance to those skills. Or the causal relationship may go both directions for different groups of students. All we can say is that the results show a definite tendency for student ratings of importance for different skills to be related to their differential degree of success in those skills.

Results for the other languages are not so clear as they are for French, but there are trends of the same general type. For example, in the German group the difference between importance ratings of Speaking and Reading predicts the relative degree of measured attainment in those skills. In Spanish the difference between importance ratings of (in particular) Writing and Reading predicts the difference between relative attainments in (in particular) Listening and Reading; that is, the student who is more oriented toward Writing (and also Listening and Speaking, to some extent) is better in oral skills than in written skills. One could give an *ad hoc* interpretation of this, but it is probably useless to do so in view of the statistical non-significance of the findings.

Relations of Vocational Plans and Other Interests to Foreign Language Skill Attainment

For 2340 "regular cases" in four languages, means and standard deviations of MLA skills test scores were computed for groups indicating various vocational plans, and compared statistically with the data for each language group as a whole. The main results are shown in Table 6.12. A remark about the constitution of the groups and the statistical computations is in order. Data are shown in the first row for "all planning to teach," i.e., all who marked the first blank in question 1 on page 4 of the Student Questionnaire (Appendix B). In the statistical computations, the scores of this group, for each language, were compared with those of all who did not mark this blank (or any of the indented spaces immediately under it). For reference, the data for the total group in each language are shown in the last row of the table. Immediately under the row (actually, two rows, one for means, one for standard deviations) for "all planning to teach" are indications as to the statistical significance of the comparisons; the footnote of the table shows the interpretation of these indications. In the next row, data are shown for those planning to teach "at the elementary school level"; this is of course a subgroup of all planning to teach. Nevertheless, statistical comparisons for this group are made not with those planning to teach at other levels but with all who did not indicate that they were planning to teach at the elementary level; e.g., for French, the 111 cases planning to teach at the elementary level were compared with the remaining (1114 - 111) = 1003 cases. Similar procedures, necessitated by the fact that there were many multiple responses, were used throughout the table. The reader who wishes to make different types of statistical comparisons may do so by using the N's, means, and S.D.'s given, but he must do so with caution because in many cases the N's overlap to some extent. For example (as may be seen in Table 6.9), there were many students who marked several levels at which they planned to teach. It should be noted, of course, that the degree of statistical significance found is partly a function of the sizes of the respective samples, and therefore it is legitimate to examine the absolute sizes of means in order to gain an idea of trends.

The main points of interest in Table 6.12 are as follows:

(1) Those planning to teach at some level, as compared to those who do not plan to teach at any level, are seldom significantly different, but when they are, they are slightly superior.

Table 6.12

Means and S.D.'s of MLA Skills Test Scores by Vocational Plans (Sexes Combined), and by Major Language

2340 "Regular" Cases in Four Languages

		FRENCH					GERMAN					RUSSIAN					SPANISH				
Group		N	L	S	R	W	N	L	S	R	W	N	L	S	R	W	N	L	S	R	W
All planning to teach	x̄	716	44.0	74.0	48.9	47.6	200	43.6	87.4	51.1	52.9	45	42.8	81.2	41.6	63.6	550	44.6	84.5	47.8	55.2
	σ		7.6	9.4	8.8	9.8		8.8	12.6	9.8	13.3		5.6	11.8	9.4	10.3		6.9	12.3	8.4	10.2
				+		+								+							+
-At elementary school level	x̄	111	44.2	74.5	47.2	46.0	12	43.5	86.2	51.4	53.8	0					53	43.7	83.7	45.8	53.2
	σ		7.0	9.0	8.7	9.4		8.2	11.6	8.8	11.9							6.9	13.9	8.7	10.8
-At secondary school level	x̄	616	43.0	73.1	47.8	46.5	156	42.6	86.1	49.4	50.8	25	42.7	78.9	40.2	62.1	492	43.9	83.8	47.0	54.2
	σ		7.6	9.3	8.6	9.9		8.8	12.4	9.3	12.9		6.0	10.4	9.4	10.6		6.8	12.2	8.2	9.9
			--		--	--				--								--		--	
-At college level	x̄	205	47.2	78.1	53.6	52.1	98	47.1	92.3	56.2	58.9	31	44.1	84.5	44.4	66.3	154	47.9	89.8	51.8	60.3
	σ		6.7	8.5	7.8	9.2		7.3	10.6	8.5	10.9		5.3	13.1	9.2	8.5		6.0	11.1	7.9	8.7
			++	++	++	++		++	++	++	++			++	+	++		++	++	++	++
All planning to use in future employment	x̄	399	45.4	73.8	49.7	48.2	121	44.9	88.5	52.2	53.7	44	42.5	78.2	41.7	62.0	299	45.2	85.2	48.8	55.7
	σ		7.3	9.5	8.4	9.3		8.0	12.0	9.6	12.7		5.7	10.7	9.5	9.8		6.8	11.6	8.4	9.8
			++		++	++		+										+	+	++	+
-In government work	x̄	236	45.8	74.7	50.4	48.4	61	45.1	89.1	51.6	54.4	36	42.2	77.5	41.6	61.6	163	45.5	85.3	48.8	55.8
	σ		7.5	9.1	8.3	9.1		8.2	11.5	9.0	12.7		5.3	11.8	9.7	9.7		6.8	11.4	8.1	9.9
			++	+	++	+												+			
-In business	x̄	183	46.1	73.9	49.7	47.7	37	43.6	86.1	49.8	49.4	0					141	45.0	84.7	48.0	54.9
	σ		6.4	8.0	7.9	8.5		8.1	13.1	8.6	11.5							6.8	11.3	8.7	10.0
			+																		
-In translating and interpreting	x̄	156	46.1	76.0	50.8	49.4	42	44.8	90.0	52.7	55.1	24	42.1	74.9	41.7	60.2	115	45.6	85.9	49.1	56.6
	σ		7.4	9.5	8.1	9.2		8.7	12.3	10.2	13.3		5.3	12.2	9.9	10.5		6.4	11.2	8.5	9.3
			++	++	++	++												+			+
-In linguistics	x̄	64	44.6	76.1	50.3	49.3	22	43.4	88.0	50.6	51.9	10	43.8	82.9	46.7	65.1	43	45.3	87.4	49.6	57.1
	σ		7.4	10.0	8.0	9.5		8.4	9.1	9.4	12.6		5.6	9.6	7.3	9.1		6.6	10.2	7.8	9.3
					+																
TOTAL GROUP	x̄	1114	44.1	73.6	48.8	47.2	324	43.4	86.8	51.0	52.1	88	42.6	78.3	41.3	62.1	814	44.5	84.0	47.8	54.6
	σ		7.6	9.2	8.6	9.5		8.6	12.8	9.8	13.3		6.0	12.7	9.6	10.9		6.8	12.0	8.4	10.1

++ Significantly ($p < .01$) superior to remainder of total group
+ " ($p < .05$) " " " " " "
- " ($p < .05$) inferior " " " " " "
-- " ($p < .01$) " " " " " "

(2) Those planning to teach at the elementary school level are generally representative of the total group. (The N's, of course, are quite small.)

(3) Those planning to teach at the secondary school level tend to be significantly inferior to those without such plans. This is particularly true for the French and Spanish groups. Nevertheless, the absolute differences are not large.

(4) Those planning to teach at the college level are strikingly superior to the remainder of the group. This is true in all languages. In fact, the superiority of the groups oriented to college teaching is the only really "stand-out" phenomenon in the whole table. One may take some comfort in this fact, for the sake of the colleges and universities, but on the other hand the lower levels of skill exhibited by those planning to teach at elementary and secondary school levels leave much to be desired. As remarked in Chapter IV, the overall skill levels of the CFLT sample in Listening and Speaking are generally at the S-2 or S-2+ levels on the FSI scale, i.e., at "Limited Working Proficiency" or a little higher. Even the college-teaching-oriented groups are not much better on this scale, on the average; only in German and in Spanish do these groups generally attain S-3, "Minimum Professional Proficiency," on the FSI scale.

(5) Students who indicated that they planned to use FL skills in some kind of future employment other than teaching tend to be slightly superior, on the average, to the remainder of the group. The differences are generally significant only in the French and Spanish groups, however, and in any case the differences are not large. One may interpret this result to mean that students who have definite vocational goals other than teaching tend to take their foreign language training more seriously than students who do not have such goals.

(6) Students oriented to careers in some kind of government work (foreign service, United Nations, A. I. D., etc.), or in translating and interpreting, are somewhat superior, in general, to students otherwise oriented.

(7) Students oriented to careers in business or in linguistics are generally representative of the total group. The finding for students thinking of careers in linguistics is somewhat counter to expectation; perhaps the students concerned do not have a good idea of what linguistics is.

Data were also tabulated for students who marked reasons other than vocational as prompting them to choose a foreign language major. A great many of these students had also marked vocational reasons (teaching and other future employment) for this choice, since they were asked to check as many reasons as applied. The results are shown in Table 6.13. Because of multiple responses, the same type of statistical comparison was used as in Table 6.12. For reference, the first row gives the data for the total sample, identical to those shown in the last row of Table 6.12. The major observations to be made from Table 6.13 are as follows:

(1) None of the groups indicating any of the reasons for choosing a foreign language major listed in the table are especially different, in terms of absolute scores, from the total group. Some of the differences are statistically significant, especially in French, but they do not appear to be of much practical significance.

(2) Over all languages, statistically significant differences are most likely to appear for groups indicating that an important reason for choosing a foreign language major was that they were "interested in intercultural understanding and communication." Such groups generally had slightly higher than average scores on the MLA skills tests.

(3) Slight superiority tended also to occur for groups indicating that they expected to use their foreign language skills in travel or study abroad, and for groups claiming they majored in a foreign language primarily because of interest in its literary and cultural aspects.

Table 6.13

Means and S.D.'s of MLA Skills Test Scores, by Reasons Other than Vocations for Choosing a Foreign Language Major, and by Major Language
2340 "Regular Cases" in Four Languages

Group	FRENCH					GERMAN					RUSSIAN					SPANISH				
	N	L	S	R	W	N	L	S	R	W	N	L	S	R	W	N	L	S	R	W
Total group	1114	44.1	73.6	48.8	47.2	324	43.2	86.8	51.0	52.1	88	42.6	78.3	41.3	62.1	814	44.5	84.0	47.8	54.6
		7.6	9.2	8.6	9.5		8.6	12.8	9.8	13.3		6.0	12.7	9.6	10.9		6.8	12.0	8.4	10.1
Expect to use in travel or study abroad	815	44.8	74.3	49.6	48.0	235	44.1	87.9	51.5	53.0	50	42.3	77.9	40.6	62.3	579	44.7	84.2	48.0	55.0
		7.4	8.9	8.2	9.2		8.3	11.3	9.2	12.6		5.8	13.0	8.5	10.4					
		++	++	++	++		+	++		+										
Majored mainly for interest and enjoyment	703	44.3	73.9	49.1	47.6	198	43.2	86.1	50.9	52.4	50	42.6	77.0	41.0	62.3	472	44.7	84.3	47.8	54.9
		7.5	9.0	8.4	9.5		8.5	12.4	9.5	13.4		5.6	11.5	8.4	10.6		6.6	11.3	8.1	10.0
					+															
Majored because of particular interest in literary and cultural aspects	583	45.1	74.6	50.4	48.7	172	44.8	88.3	52.7	53.6	41	42.3	76.8	42.5	62.1	370	44.9	85.0	48.6	55.6
		7.4	9.4	8.4	9.6		8.5	12.4	9.7	13.4		6.0	13.1	9.8	10.9		6.7	12.3	8.3	9.9
		++	++	++	++		++	+	++	+								+	+	++
Interested in intercultural understanding and communication	651	44.6	74.3	49.2	47.7	166	44.8	88.6	51.9	53.8	51	42.0	76.6	40.9	61.1	496	45.1	84.8	48.3	55.3
		7.5	9.0	8.6	9.7		8.4	12.0	9.8	13.3		5.7	12.0	8.8	10.1		6.6	11.9	8.2	10.0
		+	++		+		++	++		+							++	+	+	+
No particular reason other than the need for a major	23	45.2	76.0	48.9	47.1	13	39.1	84.7	49.5	49.1	-					19	45.2	86.6	48.1	53.7
		7.1	8.0	9.0	10.3		10.4	11.2	10.7	14.8							7.1	14.0	10.1	10.6

++ significantly (p <.01) superior to remainder of group

+ significantly (p <.05) superior to remainder of group

(4) Groups indicating that they majored in a foreign language "mainly for interest and enjoyment," or simply because they "needed a major," were representative of the total group. One might perhaps have expected that these groups would be slightly inferior to the total group in foreign language skills, but such was not the case.

5. Major Dimensions of a "Typology" of FL Majors

This section presents an attempt to use advanced statistical techniques to develop a "typology" of foreign language majors. Essentially, the technique involves finding clusters of questionnaire items that are answered in more or less the same pattern, either in one direction or its opposite. Because the procedure typically involves applying factor analysis to questionnaire items, it has been called the "QFA" technique. It has been described by Carroll et al., (1966, Appendix C), and its application to the present problem was developed by Fannie A. Handrick.

For the national sample, the technique was applied to selected questionnaire responses of 2571 students. These were all the students in the sample (including students in all five languages) who answered all questionnaire items involved in the analysis. Both "regular" and "odd" cases were included (see Chapter III, p.39).

The questionnaire items selected for study were item 5 on page 1, and items 1-11 on page 4, of the Questionnaire for Foreign Language Majors. Responses to item 5 were recoded so that they could be interpreted as "yes" or "no" answers to the question, "Have you studied any foreign language other than your major language?"

> Technical note. As a binary item with two possible answers this item contributed, in effect, two variables to the QFA covariance matrix. Each of the 18 response positions in item 1, page 4, was regarded as a binary item also; these 18 positions contributed therefore 36 variables to the QFA covariance matrix. Each of the 35 remaining response positions on page 4 (items 2 to 11) was regarded as a position on a scale and therefore contributed one variable to the QFA matrix, except that an extra "dummy" response position was created to represent non-response to item 8 (tantamount to having responded yes to either question 6 or 7). Thus, the QFA matrix contained 74 variables in all. The complete covariance matrix, generated by considering each variable as containing 1's or 0's depending upon whether that variable was coded, was submitted to a principal axes factor analysis computation routine. It yielded two latent roots greater than unity, and 43 positive latent roots in all; the remaining 31 roots were zero, corresponding to the statistical redundancy introduced by the use of non-independent responses. (For example, a student who marked "yes" in a two-choice answer would obviously not also mark "no.") Although only two latent roots were greater than unity, at least 6 of the dimensions appeared to be interpretable after orthogonal rotation.

Table 6.14 shows factor loadings for selected variables on axes rotated orthogonally from the principal axis factors. Redundant variables, and variables with low communalities, were excluded from the table.

The six factors found represent independent ways in which the FL majors in the sample varied.

Probably the most important way in which they vary is whether they plan to teach, or plan to use their foreign language skills in business or professional work. This dimension is reflected in Factor A, whose larger factor loadings are as follows:

	Factor A loading
Plans use of FL in business or professional work	.41
Plans use in government work	.29
Plans use in business	.20
Plans use as a translator or interpreter	.18

- - - - - -

Table 6.14

Factor Loadings for Selected Questionnaire Responses

(Orthogonal Rotated Axes)

Variable	Factor A	 B	 C	 D	 E	 F
Has had other mod. FL's	.03	.02	-.03	.04	.05	.12
Plans to teach at some level	-.18	.05	.34	.08	.02	.04
Plans to teach at sec. sch. level	-.21	.09	.40	-.07	.00	.04
Plans to get teaching license	-.22	.03	.37	-.07	-.01	.01
Plans to teach, college level	-.01	.05	.01	.24	.05	.00
Plans use in business or prof. work	.41	-.02	.02	.00	.11	.03
" " " business	.20	-.02	-.01	-.02	.05	.00
" " " government work	.29	-.02	.02	-.01	.08	.03
" " as translator or interpreter	.18	.00	.03	.03	.09	.02
Plans use in travel or study	.06	.04	.08	-.10	.26	-.03
Majored for interest and enjoyment	-.12	.01	-.09	-.24	.22	.02
Interested in lit. and civilization	-.13	-.02	-.09	.11	.34	-.01
Interested in international understanding	.02	.05	.03	-.05	.34	-.04
Listening is of great importance	.02	.04	.07	.03	.07	-.03
Reading " " " "	-.06	.01	.00	.17	.08	.02
Writing " " " "	-.06	.04	.07	.23	.09	-.03
Has taken informal courses	.02	.24	-.05	.01	.06	.00
Took Latin or Greek in high school	.00	.05	-.05	.00	.02	.45
Had summer or school year abroad	.06	.43	-.06	-.03	.01	-.01
No use of FL at home	-.01	-.03	-.03	-.12	.00	.09
Moderate outside use of the FL	.01	-.01	.02	-.03	.10	-.01
Has read at least 3 books in FL	.06	.16	-.05	.22	.09	-.04

Plans to teach at some level	-.18
Plans to teach at secondary school	-.21
Plans to get a teaching license	-.22

The negative loadings reflect the fact that those who state that they plan to use their foreign language skills in business or professional work show a considerable tendency not to indicate any plan to teach.

Factors C and D indicate that planning to teach at the secondary school level is largely independent of planning to teach at the college level; that is, the amount of overlap between those planning to teach at the secondary school and those planning to teach at the college level is only about what would be expected by chance. Planning to teach at the college level is associated with several other variables loaded on Factor D.

	Factor C loading
Plans to teach at the secondary school level	.40
Plans to get a teaching certificate	.37
Plans to teach (at some level)	.34

	Factor D loading
Plans to teach at the college level	.24
Writing skill is thought to be of great importance	.23
Has read at least 3 books in the FL	.22
Reading skill is thought to be of great importance	.17
- - - - - - - - - - - - - - - -	
Majored primarily for interest and enjoyment	-.24

The negative loading in the case of the last variable shows that those who plan to teach at the college level have a considerable tendency not to indicate that they choose a FL major primarily for "interest and enjoyment"; that is, their plans for college teaching are of a serious nature.

Factor B is a dimension that differentiates those who have studied abroad from those who have not; the loadings also indicate that those who have studied abroad are quite likely to have taken informal courses in their major language and to have done considerable reading.

	Factor B loading
Has had a summer or year of study abroad	.43
Has taken informal courses in the FL	.24
Has read at least 3 books in the FL	.16

Factor E seems to be an "interest" factor, differentiating those who are attracted by the intrinsic benefits of foreign language study from those who do not seem to care about these things:

	Factor E loading
Interested in the literature and civilization of the foreign country	.34
Interested in international understanding and communication	.34
Plans use of the FL in travel or study abroad	.26
Majored primarily for interest and enjoyment	.22

Finally, Factor F seems to represent simply the dimension of whether the student has studied Latin, Greek, and other languages besides his major language:

	Factor F loading
Took Latin or Greek in high school	.45
Has had other modern foreign languages besides his major language	.12

Chapter VII

EXPERIENCE AND TRAINING FACTORS IN FOREIGN LANGUAGE SKILL ATTAINMENT

1. Introduction

In this chapter we report the results of analyses in which the relations between tested foreign language attainment and various experiential and training variables were explored. Many of these variables represent kinds of influences to which the students had been exposed, whether by their own choice or by the force of circumstances. It will be tempting to infer, whenever a strong relationship between one of these variables and measured foreign language skill appears, that the "influence" in question was a causal one. Often such an inference will be very strong, particularly when it is supported by a network of related findings all pointing in the same direction. It must be remembered, however, that this study was of a purely descriptive nature and can only claim to discover interesting and suggestive relationships. The causal basis of these relationships could only be investigated properly by studies in which the variables are actually manipulated over time. For example, one of the variables to be discussed below is the extent to which the student has had the opportunity to go abroad, either as a student or as a tourist, to the country where his major foreign language is spoken. We find that students who have been abroad are better in foreign language skills than those who have not been abroad. It is perhaps reasonable to infer, from common sense and from common observation, that experience in foreign countries actually has a causal influence on the improvement of foreign language skills. On the other hand, we do not know anything about the foreign language attainments of the students _before_ they went abroad; it is conceivable that students who went abroad were the better language students before they went abroad, and that the experience abroad really had little or nothing to do with their attainments. As a matter of fact, many institutions follow the practice of selecting only their better students for admission to programs of study abroad. Also, the prospect of even a short trip abroad may stimulate the student to improve his skills markedly even before he gets on the plane or ship. Therefore, from the finding that there is a relationship between measured foreign language skill and whether the student has had a tour abroad it does not _inevitably_ follow that experience abroad causes his language skills to increase. The descriptive findings in this chapter must be taken as merely suggestive of _possible_ causal relationships. If the language used in this chapter occasionally seems to imply causal interpretations, it is only because it is sometimes difficult to find words which are sufficiently neutral in this respect.

The design of this chapter is to consider the several experiential and training factors one by one, or at most in pairs. Chapter VIII will present more elaborate statistical analyses in which large numbers of variables are considered together.

2. Time of Beginning Foreign Language Study

Our sample was composed of individuals graduating in the class of 1965. On the assumption that the typical individual went through a normal course of schooling without skipping or repeating grades, he would have entered the first grade in 1949, junior high school (Grade 7) in 1955, senior high school (Grade 10) in 1958, and college in 1961. A number of public schools, e.g. those of Cleveland, Ohio, as well as many private schools, have had foreign language programs in elementary schools for several decades, and it is possible that some of the sample had been to such schools in the early grades. Around 1953, when our sample was typically in grade 5, the "FLES" (_F_oreign _L_anguage in _E_lementary _S_chool) movement, whereby increasing numbers of schools instituted foreign language programs in pre-junior-high grades, had started to gain momentum, and it is not unreasonable to expect that many of the sample had had the opportunity to start foreign language study at that time.

On pages 2 and 3 of the Questionnaire for Foreign Language Majors, the students were asked to list all courses in their major foreign language they had taken from grade school up through the time of testing. On the basis of this list, they were classified in terms of whether they began the study of their foreign language in "grade school," in high school, or in college. (Students who listed no courses at the grade school or high school level and no beginning language course at the college level were relegated to an "odd type" category and were excluded from further analysis.) Table 7.1 shows the numbers and percentages of each category in each language, for all "regular" cases in all languages except Italian.

Table 7.1

Numbers and Percentages of "Regular" Cases,

by Time of Beginning FL Study, by Language (Excluding Italian)

(All cases complete on Four MLA Skills Tests)

	LANGUAGE									
	French		German		Russian		Spanish		Total	
Time of Beginning	N	%	N	%	N	%	N	%	N	%
"Grade school"	148	13.3	7	2.1	2	2.2	96	11.7	253	10.8
High school	785	70.4	147	44.0	22	24.7	557	68.3	1511	64.4
College	182	16.3	170	50.9	65	73.1	163	20.0	580	24.8
Total	1115	100.0	324	100.0	89	100.0	816	100.0	2344	100.0

There are no surprises in this table, except possibly the rather substantial numbers of cases who reported courses at the grade school level in French and Spanish: 13.3% in French and 11.7% in Spanish. Most students of French and Spanish, however, started at the high school level, and a slight majority of students in German and Russian started at the college level.

Next, the mean MLA skills test scores for the different groups were computed. For this analysis, grade school and high school starters were combined in the case of German and Russian because of the small numbers of students starting in the elementary school. The results are shown in Table 7.2. The differences were tested for statistical significance by one-way analysis of variance.

In French and Spanish, the two languages which substantial numbers of students started in grade school, there are highly significant differences among the categories. In all tests, those who started at grade school were superior to those who started in high school, and these in turn were better than those who started in college. In French, the successive differences are approximately equal; that is, the differences between grade school starters and high school starters and between high school starters and college starters are similar. In Spanish, the spread between the grade school and high school starters is wider than that between the high school and college starters.

The trend is in the same direction in Russian, although the differences show up with less statistical significance because of the small numbers of cases, and of course the

Table 7.2

Means and S.D.'s of MLA Skills Test Scores for Groups Starting FL Study at Different Educational Levels

(All "Regular" Cases Complete on MLA Skills Test Scores)

Time of		Listening		Speaking		Reading		Writing	
Beginning	N	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ
				FRENCH					
Grade School	148	47.47	6.46	78.09	8.50	51.64	7.75	51.40	8.61
High School	785	44.21	7.54	73.48	8.95	48.85	8.51	47.32	9.36
College	182	40.99	7.63	70.19	9.58	46.30	8.80	42.99	9.35
Total	1115	44.11	7.62	73.55	9.25	48.81	8.59	47.15	9.54
F		31.2**		31.5**		16.2**		33.9**	
				GERMAN					
Gr.Sch.& H.S.	154	42.80	8.95	86.53	13.64	50.64	10.00	51.14	13.76
College	170	44.02	8.26	86.95	11.99	51.38	9.51	52.94	12.86
Total	324	43.44	8.62	86.75	12.80	51.03	9.75	52.09	13.32
F		0.6		0.1		0.5		1.5	
				RUSSIAN					
Gr.Sch.& H.S.	24	44.00	5.73	78.79	9.83	44.63	9.06	64.42	10.32
College	65	41.98	5.96	78.11	13.48	40.00	9.48	61.20	10.89
Total	89	42.56	5.99	78.34	12.67	41.28	9.64	62.10	10.89
F		2.0*		0.0		4.2**		1.5	
				SPANISH					
Grade School	96	48.60	5.31	89.43	13.91	52.83	7.71	60.18	9.02
High School	557	44.29	6.76	83.60	11.31	47.44	8.43	54.32	10.04
College	163	42.66	6.65	82.15	12.23	46.21	7.58	52.37	9.84
Total	816	44.47	6.79	84.01	12.00	47.82	8.37	54.60	10.11
F		25.2**		12.4**		21.7**		19.6**	

** $p <.01$
* $p <.05$

grade school and high school starters were combined.

The trends are reversed in German, but the differences in means nowhere reach statistical significance despite the sizable numbers of cases. Again, the small number of grade school starters was combined in the analysis with high school starters.

No immediate explanation of these discrepant trends in German is available.

The clearly significant trends in French and Spanish, with the parallel trend in Russian, call for careful examination. The simplest qnd most conservative explanation of the trend is that the attainment of skill in a foreign language is a function of the amount of time spent in its study; it may be assumed that most of the students involved in the investigation had studied the foreign language more or less continuously since the time of first starting that study. It is not surprising, therefore, that those who started earliest had, on the average, achieved the most by the time of college graduation.

Some readers may be tempted to interpret these findings as telling justification for the FLES movement. Such a conclusion would be premature. It would be justification for the teaching of foreign languages in the elementary school only if it could also be shown that it is critical or necessary that the start of language study must be in the grade school. This could hardly be the case, for many students found in our sample who started in high school or college did as well as many students who started earlier. It must be pointed out, also, that the small number of people who appear in our data as having started their foreign language in grade school probably constitute only a small fraction of all those who started foreign language in grade school at the same time. Any conclusion about the value of having started a foreign language in the elementary school would have to take account of what happened to the rest of the students in the age cohort who started a foreign language in grade school. Our data therefore cannot be used to give unqualified support to the FLES movement. The conclusion that does seem to emerge from the data is that for those students who were enabled to start French or Spanish in the elementary school and who liked the language well enough to impel them to continue with it to the point of graduating from college with a language major, their start in elementary school gave them a distinct advantage, on the average, over those who started later.

In any case, the data presented thus far merely show the overall differences among those who started at various educational levels. In order to make stronger inferences about the effects of having started early, we must rule out other possible sources of the differences observed, such as differences in language aptitude or in other kinds of educational advantages. Such possibilities will be discussed at a later point in this report.

It might be asked whether Time of Beginning is in any way related to language aptitude. On the one hand, there might be selective factors whereby students beginning language study early (and continuing) might be students with greater language aptitude than those starting later. On the other hand, early starting might have the effect of enhancing language aptitude in some manner that would be reflected in language aptitude test scores. The data available in this investigation cannot give reliable information that would point definitely to one rather than the other of these possibilities; questions concerning explanation of any relationships found between language aptitude and Time of Beginning FL Study could be answered only by studies specifically designed to answer these questions. Nevertheless, the data assembled here may be of some interest; they are shown in Table 7.3. The differences in language aptitude among the various starting groups are in every case quite small in absolute magnitude. Only some of them are statistically significant, and they do not form a consistent pattern. For example, in French, early starters are significantly superior on MLAT-3, a test of "phonetic coding ability," but are significantly inferior on MLAT-4, a test of "grammatical sensitivity," and show no significant trend on MLAT-5. Nevertheless, the slightly inferior performance of college starters on MLAT-3 holds up across languages--as a significant effect in the Spanish group, and as a non-significant trend in the German and Russian groups.

Table 7.3

Mean MLAT Scores for Groups Starting their Major Foreign Language at Different Times

Time of beginning study of major language	N	Mean Scores on Sections of the MLAT		
		Spelling Clues MLAT-3	Words in Sentences MLAT-4	Paired-Associates MLAT-5
		FRENCH MAJORS		
Grade school	146	28.45	30.36	18.99
High school	784	27.62	32.48	19.48
College	179	25.75	32.26	19.65
Total	1109	27.43	32.17	19.69
F-ratio		3.6*	6.8**	1.8
		GERMAN MAJORS		
Grade school	7	25.59	32.33	19.49
High school	131			
College	151	24.87	32.88	19.68
Total	289	25.21	32.62	19.58
F-ratio		0.4	0.5	0.1
		RUSSIAN MAJORS		
Grade school	2	31.00	35.00	21.00
High school	20			
College	59	28.92	30.92	19.97
Total	81	29.48	32.02	20.25
F-ratio		0.5	4.1*	0.7
		SPANISH MAJORS		
Grade school	99	27.95	29.09	18.83
High school	544	24.82	30.42	19.06
College	163	23.05	30.08	17.94
Total	806	24.84	30.19	18.80
F-ratio		7.7**	1.4	2.6

**$p < .01$; *$p < .05$

On the whole, the small magnitude of the MLAT differences across groups starting at different times appears to signify that MLAT scores are largely independent of time of starting the major foreign language. In any case, it should be remembered that groups starting their major foreign language late are not necessarily without early language training experiences in other languages besides the one in which they are majoring.

3. Instructional variables

Besides data on the time of beginning FL study, considerable information was available as to the numbers and types of courses the students had had, as well as the specific kinds of instructional experiences they had had.

It was possible, first of all, to analyze more in detail the numbers and kinds of courses that students in different starting groups had had at three educational levels. Table 7.4 shows these data along with mean MLA skills test scores. What is clear from the table is that the achievement of the students on the tests can be interpreted as associated with the amount of time they had spent on language study. The substantial numbers of students who started French or Spanish in elementary school were likely to take more courses in high school; by the time they arrived in college they evidently did not have to take as many purely "language" courses as those who started in high school or college, and they tended therefore to take more foreign language literature courses. As a result, we may presume, they were able to achieve higher scores on the MLA skills tests. (The data for the small numbers of elementary school starters in German and Russian are generally consistent with the findings for French and Spanish.) Likewise, those who started in high school had to take fewer college language courses and were enabled to take more literature courses than those who started in college; their MLA test scores were correspondingly higher. This latter statement is true, at least, for students in French and Spanish; the smaller number of college literature courses taken by high school starters in German and Russian may have been restricted by the limited availability of advanced college literature courses in those languages.

It would be desirable to define as a variable the total number of semesters the student had studied a foreign language, but it was not possible to do so from the data, not only because data were unavailable but also because it would be impossible _a priori_ to equate semesters or hours at the various levels. (Students reported number of credit hours at the college level, but the number of credit hours per course varied widely.) Traditionally, foreign language teachers have considered one year of a high school course to be equivalent to one semester of a college course. If we accept that assumption, and assume further that _three_ semesters of work at the grade school level is equivalent to one semester at the college level, and that the typical college semester has three credit hours, we can compute the total number of "virtual college semesters" each starting group had; these data are shown in one of the columns of Table 7.4. The rank orders of the resulting figures, across starting groups, correspond closely to those of the MLA skills test means. It would appear that perhaps the reason why those who started German in college did slightly better on the tests than those who started in high school is that they had slightly more total exposure to the language in terms of "virtual college semesters." It is also noteworthy that the total exposure figures are greatest for the French group as a whole and least for the Russian group as a whole; this may explain the fact that the Russian group did relatively poorly on the tests when evaluated against FSI equivalents.

There exists a somewhat better procedure for showing the relative values of work at different educational levels, namely, linear regression analysis. Since this technique necessitates fairly large numbers of cases to show reliable findings, it was applied only for the French and Spanish groups, with results shown in Tables 7.5 and 7.6. This technique determines the optimal linear weights by which the predictor variables may be multiplied to yield a prediction of a dependent variable. Thus, if we examine the "b-weights" in the table for the French group (Table 7.5), we can multiply the number of semesters in grade school by .3180, the number of semesters in high school by .4698, the number of hours of college language by .0797, and the number of hours of college literature by

Table 7.4

Mean Semesters or Course Hours Taken by Groups Starting at Different Times, with Mean MLA Skills Test Scores

Group	N	Sem. Grade School	Sem. H.S.	Hrs. College "Grammar" Courses	Hrs. College Lit. Courses	Total "Virtual College Semesters"	MLA Means			
							L	S	R	W
FRENCH										
Began in grade sch.	139	4.99	6.63	13.37	20.84	16.38	47.3	77.9	51.4	51.2
Began high school	733	0.00	5.29	17.58	18.28	14.60	44.2	73.5	48.9	47.3
Began college	167	0.00	0.00	26.85	14.01	13.62	41.2	70.1	46.6	43.3
Total	1039	0.67	4.62	18.50	17.94	14.68	44.1	73.5	48.9	47.2
GERMAN										
Began in grade sch.	7	10.86	6.57	10.00	18.14	16.28	55.7	105.3	62.7	66.1
Began high school	131	0.00	4.38	15.19	16.95	12.90	42.0	85.8	50.1	50.5
Began college	151	0.00	0.00	22.86	17.12	13.32	43.7	86.3	51.0	52.3
Total	289	0.26	2.14	19.07	17.07	13.20	43.2	86.5	50.9	51.9
RUSSIAN										
Began in grade sch.	2	1.60	3.50	6.41	15.44	9.57	42.4	72.7	45.5	66.5
Began high school	20	0.00	3.25	23.53	9.81	12.74	44.4	79.3	45.6	65.0
Began college	58	0.00	0.00	26.81	10.14	12.32	42.4	78.7	40.5	61.5
Total	80	0.04	0.90	25.48	10.19	12.35	42.9	78.7	41.9	62.5
SPANISH										
Began in grade sch.	91	4.46	5.01	13.16	19.49	14.87	48.8	89.9	53.0	60.1
Began high school	518	0.00	5.15	15.06	17.74	13.51	44.3	83.5	47.5	54.3
Began college	154	0.00	0.00	25.17	14.84	13.34	42.6	81.9	46.3	52.2
Total	763	0.53	4.10	16.88	17.36	13.64	44.5	84.0	47.9	54.6

Table 7.5

Regression System for Predicting MLA Skills Test Scores from Amount of Study at Different Educational Levels

FRENCH

A. Basic Data for Total Group (N = 1039)

		Predictor Variables					Criterion Variables (MLA Tests)			
		1	2	3	4		Listen.	Speak.	Read.	Write
Mean:		0.67	4.62	18.50	17.94		44.14	73.52	48.89	47.18
S.D.:		2.42	2.84	9.99	8.04		7.59	9.67	8.51	9.53
		Intercorrelations					Beta-weights			
No.sem.grade sch.	1	1.00	.21	-.17	.13	1	.1015	.1262	.0847	.1002
No.sem. high sch.	2	.21	1.00	-.42	.32	2	.1758	.1984	.0954	.1484
Hrs. college lang. courses	3	-.17	-.42	1.00	- 21	3	.1046	.1217	.0158	.0354
Hrs. college lit. courses	4	.13	.32	-.21	1.00	4	.3025	.1756	.2918	.2669
						Multiple R:	.4010	.3241	.3544	.3654
		Correlations with Criteria					b-weights			
	L	.16	.25	-.05	.35	1	.3180	.4830	.2978	.3945
	S	.17	.23	-.02	.23	2	.4698	.6480	.2859	.4984
	R	.14	.20	-.10	.33	3	.0797	.1131	.0136	.0334
	W	.16	.24	-.10	.32	4	.2854	.2021	.3089	.3164
						Intercept:	35.1620	64.4846	41.5764	38.3190

B. Prediction of Criterion Means for Groups Starting at Different Times

		Predicted Values of Criterion Means (Actual Values in Parentheses)			
Began in Grade School	139	46.88	76.91	51.58	50.63
		(47.34)	(77.88)	(51.45)	(51.23)
Began in High School	733	44.27	73.60	48.97	47.33
		(44.20)	(73.47)	(48.93)	(47.31)
Began in College	167	41.30	70.35	46.27	43.65
		(41.22)	(70.14)	(46.59)	(43.26)

Table 7.6

Regression System for Predicting MLA Skills Test Scores from Amount of Study at Different Educational Levels

SPANISH

A. Basic Data for Total Group (N = 763)

		Predictor Variables					Criterion Variables (MLA Tests)			
		1	2	3	4		Listen.	Speak.	Read.	Write
Mean:		0.53	4.10	16.38	17.36		44.50	83.95	47.88	54.59
S.D.:		1.91	2.78	9.39	7.94		6.32	12.07	8.39	10.19
		Intercorrelations					Beta-weights			
No.sem.grade sch.	1	1.00	.01	-.12	.05	1	.2115	.2509	.1936	.1712
No.sem. high sch.	2	.01	1.00	-.43	.21	2	.0860	.0111	.0055	.0853
Hrs. college lang. courses	3	-.12	-.43	1.00	-.21	3	-.0277	-.0153	-.1085	-.0348
Hrs. college lit. courses	4	.05	.21	-.21	1.00	4	.2856	.1419	.2664	.2763
						Multiple R:	.3974	.2995	.3791	.3702
		Correlations with Criteria					b-weights			
	L	.23	.16	-.15	.32	1	.7550	1.5860	.8507	.9130
	S	.26	.05	-.08	.16	2	.2107	.0483	.0168	.3128
	R	.22	.11	-.19	.30	3	-.0205	-.0193	-.0973	-.0377
	W	.19	.16	-.15	.31	4	.2455	.2160	2.810	.3546
						Intercept:	39.3201	79.4874	44.1245	47.3042

B. Prediction of Criterion Means for Groups Starting at Different Times

		Predicted Values of Criterion Means (Actual Values in Parentheses)			
Group	N	L	S	R	W
Began in Grade School	91	48.26	90.76	52.20	59.36
		(48.79)	(89.86)	(52.95)	(60.08)
Began in High School	518	44.45	83.28	47.73	54.60
		(44.31)	(83.52)	(47.46)	(54.32)
Began in College	154	42.45	82.21	45.85	51.62
		(42.62)	(81.89)	(46.30)	(52.23)

.2854, then add 35.1620, to yield the estimated MLA Listening score for a given student. (These weights are to be contrasted with weights of .3333, .5000, .3333, and .3333, respectively, implied by the weighting system described above whereby the number of "virtual college semesters" is computed.) It would appear that for the total French group, hours of college language courses are of much less significance than is traditionally thought: they are of relatively little weight, certainly, for those who have already had a solid introduction to the language in grade school or high school. On the other hand, college literature courses are shown to be influential in promoting language competence.

Tables 7.5 and 7.6 show, in general, that a considerable proportion of the variance of MLA skills test scores is associated with the amount of time students have spent in language study. The mean test scores of groups starting at different educational levels are quite accurately predictable. Listening test scores are on the whole more predictable in this way than scores on the other tests, a finding that suggests that the Listening tests are the most valid measures of language competence.

The negative weights of college language courses for the Spanish group suggest that students arriving at college with previous study who still had to take college language courses may have been the poorer students. (These courses would probably not have negative weights, however, if data for students starting in college had been analyzed separately.)

Instructional Procedures

The course information collected from the students touched various aspects of instructional procedures. We have already seen, in Chapter V, that institutions differ markedly in some of these variables when the institutions are classified by size and type (private vs. public). Information was available on instructional procedures not only in college but also in high school, insofar as the students could be trusted to remember those aspects of their high school courses. The data are sufficiently consistent over languages, however, to justify considerable confidence in them.

We shall focus attention on just four of these variables:

(1) The degree to which the foreign language was used in the classroom by the teacher.

(2) The degree to which the student was required to use the FL.

(3) The degree to which a language laboratory was used, and if so, the extent to which it was an integral part of the course.

(4) The extent to which the teacher was judged to have a native or near-native accent in the foreign language.

Each of these variables was averaged over groups of courses--high school courses, and college beginning and intermediate language courses. Some results obtained for these variables are shown in Table 7.7. The first part of this table concerns groups of students who started FL study in high school, and therefore it presents data concerning instructional variables applicable at that stage. If we look at means, we notice the following about high school foreign language instruction as reported by the students:

1. Teaching was mostly in English.
2. In the main, the student's classroom language was English.
3. Generally, there was very little use of language laboratories.
4. The average teacher's ability to pronounce the FL correctly was about halfway between extremes of incompetence and the competence of a native speaker.

Table 7.7

Data for Certain Instructional Variables and Correlations with MLA Skills Tests

A. Group Starting in High School

Variable	Lang.	N	Mean	S.D.	Correlations with MLA Skills Test			
					L	S	R	W
Teacher's use of FL, h.s.	F	733	1.74	0.60	.09*	.06	.02	.06
Coding: 1=Mostly English used 3=Mostly FL used	G	131	1.61	0.64	.20*	.19*	.20*	.18*
	S	519	1.73	0.65	.12**	.13**	.13**	.14**
Student use of FL, h.s.	F	733	1.29	0.40	.09*	.07*	.04	.06
Coding: 1=Mostly English used 2=Mostly FL used	G	131	1.20	0.35	.24**	.21**	.22*	.17
	S	519	1.26	0.38	.08	.09*	.08	.14**
Language lab. importance, h.s. courses	F	733	1.22	0.44	.06	.02	.02	.03
Coding: 1=No language lab. 3=Lang.Lab. important and integral part	G	131	1.22	0.47	.08	.00	.15	.09
	S	519	1.22	0.49	.05	.02	.08	.06
Teacher's pronunciation, h.s. courses	F	733	1.95	0.70	.04	.08*	.00	.02
Coding: 1=Definite non-native accent 3=Native accent	G	131	2.22	0.74	.16	.13	.18*	.11
	S	519	2.11	0.68	.04	.05	.02	.00

B. Group starting in college

Variable	Lang.	N	Mean	S.D.	Correlations with MLA Skills Test			
					L	S	R	W
Teacher's use of FL, coll. lang. courses	F	167	2.20	0.50	.27**	.27**	.17*	.18*
Coding: 1=Mostly English used 3=Mostly FL used	G	151	2.08	0.58	.23**	.14	.14	.22**
	R	58	2.17	0.53	.07	-.01	.05	.08
	S	154	2.24	0.53	.39**	.32**	.35**	.36**
Student use of FL, college lang. courses	F	167	1.54	0.31	.21**	.19*	.13	.15
Coding: 1=Mostly English used 2=Mostly FL used	G	151	1.47	0.35	.30	.20	.20	.28
	R	58	1.61	0.31	.29*	.20	.29*	.25
	S	154	1.53	0.32	.32**	.29**	.26**	.28**
Language lab. importance, coll. lang. courses	F	167	1.88	0.54	-.03	-.03	-.03	-.13
Coding: 1=No lang. lab. 3=Lang. Lab. important	G	151	1.81	0.57	-.05	-.15	-.08	-.14
	R	58	1.94	0.56	-.02	-.02	-.03	.05
	S	154	1.72	0.58	-.11	-.05	-.03	-.14
Teacher's pronunciation, coll. lang. courses	F	167	2.56	0.47	.17	.14	.13	.15
Coding: 1=Def. non-native accent 2=Native accent	G	151	2.65	0.42	.18*	.07	.12	.17*
	R	58	2.70	0.41	-.07	-.07	-.14	-.08
	S	154	2.56	0.42	.10	.12	.12	.13

On the other hand, there was considerable variation in students' experiences. Some students had apparently had the benefit of teaching that employed much active foreign language use by both teacher and student. Although it was a little too much to hope that such variations would be associated with performance on the MLA Skills tests taken at least 4 years after the high school experience, Table 7.7 actually shows a number of significant correlations, and these are more or less consistent across languages. The teacher's use of the FL in the classroom and the student's use of it are about equally correlated with MLA skills test scores. The excellence of the teacher's pronunciation does not seem to make much difference, and variations in emphasis on language laboratory use do not correlate at all with performance on the MLA tests. (The latter finding may reflect the fact that good teaching may be done either with or without a language laboratory.)

The second part of the table concerns groups that started study of the FL in college, and the instructional variables apply to college courses only. The mean scores for the variables indicate that instructional procedures in college differ markedly from the high school, at least according to these students' reports. Much more use of the foreign language is made both by the teacher and the student, and teachers' pronunciations are more likely to be of a native or near-native character. There is considerably more emphasis on the use of language laboratories, but it should be noted that whether a student has had a language laboratory experience has little or nothing to do with his eventual attainment of skill: indeed, the trend is toward a negative (but never significant) correlation. But teacher and student use of the FL in the classroom is frequently significantly associated with tested skill.

It should be borne in mind that MLA skill scores are associated with many variables besides instructional ones, as we have seen and shall see; the fact that statistically significant correlations were obtained at all for these instructional variables does bear testimony, however, for the proposition that quality of instruction has at least something to do with student attainment. When quality of instruction, as measured by the instructional variables mentioned here, is combined appropriately with information on student aptitude, amount of exposure to the language, and other variables, it is possible to predict attainment more accurately than otherwise, as postulated by the writer in a previous publication (Carroll, 1963a).

4. Time spent abroad

On page 4 of the Questionnaire for Foreign Language Majors (Appendix B), students were asked to indicate what kind of experience they may have had in travelling or studying in the country where their major language is spoken natively. On the basis of the responses, students were classified into the following groups:

1. Students who had had a school year abroad ("a regular school year in a country where your major language is spoken natively and took a series of courses in your major language during that time")
2. Students who had had a summer of study abroad (but not a year abroad)
3. Students who had had neither a school year nor a summer of study abroad but who had travelled as a tourist in the major-language country and/or had had a program of self-study there
4. Students who had never been abroad to the major-language country.

Table 7.8 shows the numbers of such students, by language and by sex, along with mean MLA skills test scores for each sub-classification.

Substantial numbers of students had been abroad to a country where their major language was spoken: 51.5% of the French majors, 60.1% of the German majors, 22.6% of the Russian majors, and 55.3% of the Spanish majors. There was no significant difference between men and women in this respect for French and German; however, significantly more men (36.7%) than women (14.8%) had been abroad in the Russian group, while significantly

Table 7.8

Mean MLA Skills Test Scores for Groups with Different Amounts of Experience Abroad, by Language and by Sex

	N's			Means of MLA Skills Test Scores											
				Listening			Speaking			Reading			Writing		
	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
FRENCH															
Year abroad	39	204	243	50.46	50.16	50.21	79.72	79.02	79.14	55.72	53.26	53.65	54.05	52.84	53.05
Summer abroad	25	153	178	46.04	45.72	45.76	77.36	75.15	75.46	50.88	49.92	50.06	48.96	48.27	48.37
Toured abroad	23	120	143	45.78	45.06	45.17	76.96	74.02	74.49	48.13	49.31	49.12	44.39	47.79	47.24
Never abroad	87	445	532	40.07	40.66	40.57	71.38	69.96	70.19	46.43	46.13	46.18	42.85	43.39	44.14
Total	174	922	1096	44.01	44.17	44.15	74.84	73.35	73.59	49.37	48.75	48.85	46.44	47.35	47.20
F-ratio				20.9**	106**	128**	8.3**	59**	66**	9.2**	41**	50**	13.0**	45**	57**
Chi-square	0.5§ (3 d.f.)														
GERMAN															
Year abroad	28	51	79	49.96	50.39	50.24	97.82	93.49	95.03	58.61	57.71	58.03	61.29	61.12	61.18
Summer abroad	15	18	33	46.93	44.28	45.48	88.27	88.33	88.30	53.73	50.28	51.85	55.87	54.17	54.94
Toured abroad	37	42	79	46.14	45.83	45.97	90.32	91.29	90.84	53.46	52.93	53.18	55.78	54.55	55.13
Never abroad	49	78	127	36.92	37.55	37.31	79.41	78.31	78.73	45.61	45.21	45.36	44.45	43.64	43.95
Total	129	189	318	43.56	43.50	43.52	87.57	86.24	86.78	51.13	50.78	51.12	52.68	51.78	52.15
F-ratio				27**	38**	66**	21**	22**	41**	14**	26**	40**	15**	25**	41**
Chi-square	2.7 (3 d.f.)														
RUSSIAN															
Year abroad Summer abroad Toured abroad	11	8	19	45.91	46.00	45.95	87.82	84.50	86.42	43.54	45.01	44.16	67.73	66.00	67.00
Never abroad	19	46	65	40.05	42.41	41.72	75.00	76.70	76.20	38.37	41.41	40.52	58.79	61.42	60.65
Total	30	54	84	42.20	42.94	42.68	79.70	77.86	78.51	40.27	41.94	41.34	62.07	62.10	62.09
F-ratio				7.7**	§	7.9**	6.5*	§	10.5**	2.3	§	2.0	4.4*	§	5.0*
Chi-square	5.3* (1 d.f.)														
SPANISH															
Year abroad	20	101	121	50.30	50.5	50.55	96.35	90.95	91.84	57.55	54.21	54.76	64.55	62.18	62.57
Summer abroad	25	136	161	47.56	44.73	45.17	91.28	83.32	84.55	51.24	47.35	47.95	57.84	54.47	54.99
Toured abroad	55	103	158	45.24	45.63	45.49	87.82	85.75	86.47	48.20	49.35	48.95	54.78	56.15	55.67
Never abroad	98	257	355	41.61	41.82	41.76	81.54	79.67	80.18	45.16	44.96	45.02	50.67	51.67	51.39
Total	198	597	795	44.25	44.62..	44.53	86.01	83.46	84.09	48.03	47.83	47.88	54.12	54.86	54.67
F-ratio				13.6**	54**	66**	12.8**	26**	35**	15.8**	37**	50**	12.0**	33**	44**
Chi-square	21.8** (3 d.f.)														

** p<.01; *p<.05; §not computed because of small N more women (56.9%) than men (50.5%) had been abroad in the Spanish major group.

In MLA skills test scores, there were highly significant differences in every language group among groups with different amounts of experience abroad, favoring those who had been abroad, and particularly those who had been abroad for a year of study. Persons who had had only a summer abroad, or a tour, had mean test scores intermediate between those who had had a year abroad and those who had never been abroad. Since it did not seem to make much difference in these intermediate groups whether the experience abroad was a summer or a tour, these groups were combined for subsequent analyses, and the Time Abroad variable was re-coded so as to constitute three categories:

3: A year abroad

2: Summer study or a tour abroad

1: Never abroad

As stated in the opening paragraphs of this chapter, it is tempting to conclude that going abroad is an important causative variable influencing language competence, and such a conclusion accords with common experience. Indeed, the highly significant effects reported above are not likely to have occurred simply as a result of selective factors. Nevertheless, until more information is available from studies based on pre-test and post-test comparisons of groups who are and who are not sent abroad, it is impossible to state with certainty--at least on the basis of the present data--just what the effect of travel and study abroad may be. Carton and Carroll (1960) gave pretests and posttests to groups of Russian students making a six-week tour of the Soviet Union and concluded that the tour yielded distinctly increased competence, especially in speaking skills, and especially for students with at least two years' prior study of Russian, but they were not able to compare the effects of the tour with other kinds of language training experiences.

The effect of Time Abroad is independent of Time of Beginning FL Study, as shown by Table 7.9 for the Listening test scores for French, German, and Spanish groups. The F-ratios for the interaction of the two variables are, respectively, .26, .10, and .26--all insignificant. Similar results, not shown here, were obtained for the Speaking, Reading, and Writing tests. Incidentally, there are significant tendencies shown in Table 7.9 for persons who started their language study early to have had more experience abroad than those who started their language study later. In French, the percentages of students who have had at least some experience abroad are as follows: grade school starters, 59.7%; high school starters, 50.3%; college starters, 50.3%. In Spanish, the percentages are, respectively, 69.2%, 52.9%, and 50.0%. However, the trend is reversed in German: only 49.6% of the high school starters had been abroad, while 60.9% of the college starters had. The findings in French and Spanish possibly reflect economic factors whereby those having FLES are more likely to have come from families who could afford to send them abroad. The lack of significant interactions in the effects of the two variables on MLA skills test scores suggests, however, that the students who traveled abroad after FLES experience did not gain any special advantage thereby. (That is, the fact that they had had FLES experience did not make them more susceptible to improvement by foreign study than students without such FLES experience.)

Time Abroad appears to be a more influential variable than Time of Beginning FL Study, to judge from the fact that the former shows wider separation of means by subclassifications.

In none of the languages were the groups classified by amount of time abroad significantly different on any of the MLAT (language aptitude) scores. Going abroad thus does not appear to depend on language aptitude as a self-selection factor, and it does not affect MLAT scores, according to these results.

Table 7.9

Mean Listening Test Scores (and N's)
by Time of Beginning FL Study and by Time Spent Abroad

	Time Began			
	Grade School	High School	College	Total
		FRENCH		
Year Abroad	51.74 (38)	50.22 (160)	47.52 (31)	50.11 (229)
Tour or Summer Abroad	48.16 (45)	45.55 (209)	42.15 (53)	45.34 (307)
Never Abroad	43.70 (56)	40.78 (364)	38.04 (83)	40.65 (503)
Total	47.34 (139)	44.20 (733)	41.10 (167)	44.12 (1039)
		GERMAN		
Year Abroad	(Not used	50.03 (30)	50.65 (37)	50.37 (125)
Tour or Summer Abroad	in computations)	44.71 (35)	45.89 (55)	45.43 (90)
Never Abroad		36.95 (66)	37.27 (59)	37.10 (125)
Total	-	42.02 (131)	43.69 (151)	42.91 (282)
		SPANISH		
Year Abroad	52.55 (29)	50.08 (73)	47.92 (13)	50.46 (115)
Tour or Summer Abroad	48.53 (34)	45.10 (201)	44.75 (64)	45.42 (299)
Never Abroad	45.21 (28)	41.92 (244)	39.96 (77)	41.75 (349)
Total	48.79 (91)	44.30 (518)	42.62 (154)	44.50 (763)

5. Use of the Major FL at Home

Students were asked, "Do you and/or your parents speak your major language at home?" and were allowed to check either "No," "Yes, occasionally," or "Yes, frequently." The ambiguity of the question is perhaps unfortunate, because we cannot tell whether the student himself used the foreign language with his parents. The question was intended merely to reveal whether the student was exposed to use of the foreign language in his home. The means of MLA skills test scores for students answering each alternative are shown, for each language (except Italian), in Table 7.10. Highly significant differences are nearly everywhere apparent, favoring those with frequent use of the FL at home. The groups used in the analysis were "regular" cases only; students identified as native speakers of the language in question had been excluded. Those who claim "frequent" use of the FL at home either by themselves or by their parents do almost as well as students classified as native speakers (see Chapter IV). Obviously home is an excellent place to learn a foreign language if it is regularly spoken there. There are small numbers of students in our sample who had that opportunity, and substantial numbers whose homes were ones where the FL was used "occasionally." This fact must be held in mind in interpreting norms presented in Chapters IV and V.

6. Other Experiences in the Foreign Language

Students were asked, "Considering only your major language, have you taken informally (not for credit) any course, or audited any course, in that language (whether summer, night, special, or other)?" Of 1114 French "regular" cases, 29.1% indicated "yes" to this question. The figures for the other languages are: of 324 German students, 29.6% "yeses"; of 88 Russian students, 19.3%; of 814 Spanish students, 24.9%. For French, German, and Spanish students, highly significant differences were found on each of the MLA skills tests, favoring those who reported having taken informal courses. We can only guess at the nature of these courses. The fact of having taken an informal course may be an indication of special motivation on the part of the student; it does not necessarily represent an influence on the student.

Another question asked was, "Have you had occasion to use your major language in a job situation, with foreign friends, or in some other extracurricular situation?" Such opportunity was probably not independent of other experiences in the backgrounds of the students, and it is not clear whether this variable should be regarded as a predictor variable or a criterion variable, because this "extracurricular" use of a foreign language might signify either a cause or an effect. At any rate, large numbers of students answered either "moderately" or "extensively" to this question, as may be seen in Table 7.11, and there were in nearly every case highly significant differences in mean MLA skills test scores among the different response groups, always favoring those with "extensive" outside use of the language, and to a lesser degree, those with "moderate" outside use, over those reporting no outside use.

The amount of independent reading that the student had done was assessed by the question, "Independent of and in addition to reading required in courses, have you had any other reading experiences in your major language?" The responses to this question constitute another variable which cannot clearly be classified as either an antecedent or a consequent of foreign language attainment, but it was of interest to learn how much independent reading the students had done and to see whether it was associated with tested achievement. Table 7.12 reports the numbers of students in each language who indicated each of the possible answers to the question given on the questionnaire: "have not done any reading except for material in course work"; "read a few pages of material"; "read one or two books (or the equivalent)"; "and "read three or more books." The distributions of responses vary according to the sex of the respondent and also from language to language. There is a tendency for the men to have read more widely than the women. Students of French tend to do much more independent reading than students of other languages; 55.3% of them report having read three or more books. 47.8% of the German students had read three or more books, 42.1% of the Spanish students, and only 22.7% of the Russian students, who, it will be recalled, tend to do relatively worse on

Table 7.10

Mean MLA Skills Test Scores by Use of the FL at Home
("Regular" Cases Only)

			Mean MLA Skills Test Scores			
Use of FL at home	N	%	Listening	Speaking	Reading	Writing
			FRENCH			
Frequently	15	1.3	49.40	83.67	53.20	53.53
Occasionally	101	9.1	46.53	77.18	50.57	49.25
Never	998	89.6	43.78	73.03	48.56	46.84
Total	1114	100.0				
F-ratio			9.8**	18.9**	4.5*	6.4*
			GERMAN			
Frequently	14	4.3	51.71	101.64	59.71	63.07
Occasionally	49	15.1	46.14	92.29	52.18	54.57
Never	261	80.6	42.49	84.91	50.34	51.03
Total	324	100.0				
F-ratio			11.1**	18.5**	6.7**	6.6**
			RUSSIAN			
Frequently	1	1.1	52.00	104.00	50.00	71.00
Occasionally	7	8.0	44.43	83.71	45.00	65.14
Never	80	90.9	42.28	77.55	40.85	61.72
Total	88	100.0				
No test of significance was run for Russian						
			SPANISH			
Frequently	49	6.0	50.14	97.59	55.45	61.22
Occasionally	75	9.2	47.84	89.71	50.83	57.59
Never	690	84.8	43.70	82.43	46.95	53.81
Total	814	100.0				
F-ratio			33**	51**	31**	16**

**p<.01, *p<.05

Table 7.11

Amount of Use of the FL in Jobs, with Friends, Etc., by Language and by Sex

Amount of FL use outside	Men N	Men %	Women N	Women %	Total N	Total %
		FRENCH				
Extensive	31	17.5	133	14.2	164	14.7
Moderate	98	55.4	491	52.4	589	52.9
None	48	27.1	313	33.4	361	32.4
Total	177	100.0	937	100.0	1114	100.0
		GERMAN				
Extensive	28	21.1	37	19.4	65	20.1
Moderate	65	48.9	101	52.9	166	51.2
None	40	30.1	53	27.8	93	28.7
Total	133	100.1	191	100.1	324	100.0
		RUSSIAN				
Extensive	4	12.1	3	5.4	7	7.9
Moderate	11	33.3	30	54.5	41	46.6
None	18	54.5	22	40.0	40	45.4
Total	33	99.9	55	99.9	88	99.9
		SPANISH				
Extensive	40	19.8	124	20.3	164	20.1
Moderate	120	59.4	345	56.4	465	57.1
None	42	20.8	143	23.4	185	22.7
Total	202	100.0	612	100.1	814	99.9

Table 7.12

Amount of Independent Reading, by Language and by Sex, with Mean Reading Test Scores for Language Groups

Amount of independent reading	Men		Women		Total		
	N	%	N	%	N	%	Mean Reading Score
				FRENCH			
None	7	4.0	38	4.0	45	4.0	44.64
Little	17	9.6	147	15.7	164	14.7	43.42
One or two books	45	25.4	244	26.0	289	25.9	46.29
Three or more books	108	61.0	508	54.2	616	55.3	49.66
Total	177	100.0	937	99.9	1114	99.9	F=31**
				GERMAN			
None	9	6.8	7	3.7	16	4.9	43.75
Little	25	18.8	27	14.1	52	16.0	43.48
One or two books	29	21.8	72	37.7	101	31.2	47.56
Three or more books	70	52.6	85	44.5	155	47.8	56.57
Total	133	100.0	191	100.0	324	99.9	F=50**
				RUSSIAN			
None	13	39.4	21	38.2	34	38.6	36.6
Little							
One or two books	13	39.4	21	38.2	34	38.6	41.3
Three or more books	7	21.2	13	23.6	20	22.7	49.2
Total	33	100.0	55	100.0	88	99.9	F=13.6**
				SPANISH			
None	14	6.9	33	5.4	47	5.8	43.4
Little	40	19.8	149	24.3	189	23.2	43.6
One or two books	52	25.7	183	29.9	235	28.9	45.9
Three or more books	96	47.5	247	40.4	343	42.1	52.0
Total	202	99.9	612	100.0	814	100.0	F=66**

**p < .01

the tests than students of other languages, and have had relatively less exposure to formal instruction. Table 7.12 also shows the mean Reading test scores for the groups of respondents according to the amount of independent reading they have done. Students who report having done no independent reading are at about the same level of achievement as those who say they have read "only a few pages" of material in the foreign language. As the amount of independent reading increased beyond "a few pages," the mean Reading test scores increased monotonically. Students who have read three or more books attain distinctly superior Reading test scores. (Scores on the other MLA skills tests show similar trends but are not reported here.)

7. Other Foreign Languages Taken Besides Major Language

From data supplied by the students, Table 7.13 has been constructed to show the frequency with which majors in a given language had studied other modern foreign languages; the cases were classified according to the time they began study of their major language. Information was not obtained as to the time(s) the students studied the other languages. Common observation would suggest that those who started their major language in grade school would most likely have studied other languages at a later time, while those who started their major language in college are most likely to have started another language at an earlier time. There is nothing in the table to suggest that those who started their major language in grade school were any more likely than those who started in high school or college to have had study of other languages.

As one might expect, students majoring in a given Romance language are fairly likely to have studied other Romance languages. Of the French majors, 45.9% have studied Spanish, and 13.3% have studied Italian. 65.2% of the Italian majors have studied French and 17.4% Spanish. 56.9% of the Spanish majors have studied some French, and 10.8% of them have studied Italian. It may be noted that although Italian is not very popular as a major, it is nevertheless studied by substantial numbers of students, mostly majors in other Romance languages.

French is a very popular second language even for majors in German and Russian. German is particularly popular with majors in Russian, and to some extent also with majors in French. Russian is a second foreign language studied by small numbers of majors in other languages; it is most popular among students of German. Spanish is studied by appreciable numbers of German and Russian majors, along with the students of Romance languages.

Considerable numbers of all the groups indicated that they had studied other modern foreign languages besides the 5 test languages. There was a great variety of such languages and no tabulations are presented here.

There has been perennial interest in the question of whether the study of one foreign language helps in the study of another language. The present investigation was not designed to study this question with any rigor, although with careful statistical analysis the data might permit certain tentative conclusions. Table 7.14 represents a first step towards such an analysis; it shows the mean MLA skills test scores of those who have had study of other languages in comparison to the means for the complete language major groups. The comparisons were tested for statistical significance by the *t*-test of the difference in the means of those who had had the other language and of those who had not. A number of comparisons were highly significant, usually favoring those who had had study of another language besides their major. For example, French majors who had also had study of German, Italian, or Russian were without exception significantly superior on tests to those who had not studied the given other language. German majors who had had French were superior to those who had not, but German majors with Spanish study were inferior on the Listening and Reading tests to those without Spanish study. A few comparisons in the Spanish group were highly significant.

It is very difficult to interpret these results at face value. There does not seem to be any consistent pattern whereby study of related languages (in the Romance group)

Table 7.13

Other Languages Taken, By Time of Beginning Major Language (Excluding Latin and Ancient Greek)

Time of Beginning Study of Major Language	Language other than the Major Language											
	French		German		Italian		Russian		Spanish		Other	
	N	%	N	%	N	%	N	%	N	%	N	%
FRENCH MAJORS												
Grade school	(148)		31	20.9	29	19.6	17	11.5	70	47.3	53	35.8
High school	(784)		188	24.0	102	13.0	81	10.3	348	44.4	294	37.5
College	(182)		58	31.9	17	9.3	10	5.5	93	51.1	72	39.6
Total	(1114)		277	24.9	148	13.3	108	9.7	511	45.9	419	37.6
GERMAN MAJORS												
Grade school	6	85.7	(7)		0	--	0	0.0	2	28.6	4	57.1
High school	66	44.9	(147)		11	7.5	20	13.6	32	21.8	51	34.7
College	96	56.5	(170)		7	4.1	28	16.5	55	32.4	82	48.2
Total	168	51.8	(324)		18	5.5	48	14.8	89	27.5	137	42.3
ITALIAN MAJORS												
Grade school	0	--	0	--	(2)		0	--	1	50.0	0	--
High school	4	50.0	0	--	(8)		0	--	1	12.5	1	12.5
College	11	84.6	0	--	(13)		0	--	2	15.4	5	38.5
Total	15	65.2	0	--	(23)		0	--	4	17.4	6	26.1
RUSSIAN MAJORS												
Grade school	1	50.0	0	--	0	--	(2)		0	--	1	50.0
High school	14	63.6	10	45.4	0	--	(22)		3	13.6	12	54.5
College	41	64.1	26	40.6	5	7.8	(64)		17	26.6	32	50.0
Total	56	63.6	36	40.9	5	5.7	(88)		20	22.7	45	51.1
SPANISH MAJORS												
Grade school	70	72.9	13	13.5	11	11.5	4	4.2	(96)		30	31.2
High school	303	54.6	77	13.9	64	11.5	38	6.8	(555)		165	29.7
College	90	55.2	28	17.2	13	8.0	10	6.1	(163)		70	42.9
Total	463	56.9	118	14.5	88	10.8	52	6.4	(814)		265	32.6
ALL MAJORS												
Grade school (N = 255)	77	30.2	44	17.2	40	15.7	21	8.2	73	28.6	88	34.5
High school (N = 1516)	387	25.5	275	18.1	177	11.7	139	9.2	384	25.3	523	34.5
College (N = 592)	238	40.2	112	18.9	42	7.1	48	8.1	167	28.2	261	44.1
Total (N = 2363)	702	29.7	431	18.2	259	11.0	208	8.8	624	26.4	872	36.9

Table 7.14

Mean MLA Skills Test Scores for Groups with Study of Other Languages Besides their Major Languages

		Mean MLA Skills Test Scores			
	N	Listening	Speaking	Reading	Writing
			FRENCH		
Had German	277	46.03**	74.84**	50.89**	49.18**
Had Italian	148	46.52**	75.47**	51.94**	50.93**
Had Russian	108	47.21**	76.22**	53.48**	52.30**
Had Spanish	511	44.28	74.12	48.98	47.54
Had other langs.	419	44.14	72.99	49.28	47.45
All French	1114	44.11	73.55	48.81	47.15
			GERMAN		
Had French	168[1]	44.62*	88.82**	52.18*	53.79*
Had Italian	18	44.17	92.11	52.61	57.22
Had Russian	48	45.42	88.67	53.71*	56.29*
Had Spanish	89	41.39**(-)	84.57	48.78*(-)	50.30
Had other langs.	137	43.67	87.28	52.03	53.73
All German	324	43.44	86.75	51.03	52.09
			RUSSIAN		
Had French	56	42.71	78.12	41.12	62.41
Had German	37	42.62	80.35	41.97	64.22
Had Italian	5	(Not computed)			
Had Spanish	20	41.45	76.90	39.00	61.90
Had other langs.	45	42.40	78.42	40.87	62.53
All Russian	88	42.56	78.34	41.28	62.10
			SPANISH		
Had French	463[1]	45.67**	85.58	48.99	56.04**
Had German	118	44.68	83.35	49.08	55.39
Had Italian	88	45.54	85.26	50.33**	57.04*
Had Russian	52	46.04	88.27*	50.52*	59.00**
Had other langs.	265	44.54	83.17	47.80	55.13
All Spanish	814	44.47	84.01	47.82	54.60

[1] These comparisons were also analyzed with covariance adjustment for language aptitude (MLAT total); the differences remained significant.

**p < .01, *p < .05 for comparison with means for those not having had study of the language in question. A minus (-) attached calls attention to the fact that the mean is lower than that of the comparison group.

is more associated with higher test scores than study of languages that are less related (Romance vs. Germanic vs. Slavic). It is possible that the comparisons reflect relative amounts of time spent on the respective languages, differences in motivation for language study, etc. For example, the lower scores of German majors who have had Spanish might reflect a lesser amount of study of German due to increased time devoted to Spanish study. Although (as mentioned above) the data of the present investigation could afford further analyses, it has not been considered worthwhile to pursue these analyses for the present report, partly because it was felt that even these further analyses would lead to more questions than they could answer, in view of the lack of data on when and how long the other languages had been studied.

Somewhat more effort, however, has been expended on trying to analyze the effect of prior study of Latin and Greek on performance on the MLA tests, since the questionnaire provided specific data on when these languages had been studied (high school or college). It was realized that any analysis should take account of not only when the study of Latin (or ancient Greek) was taken up but also the time of beginning the major language. Further, language aptitude would have to be controlled, at least statistically. Therefore, in each of the three languages in which there were sufficient numbers of cases (French, German, and Spanish), a two-way analysis of variance was performed with Time of Beginning the major language and Time of Study of Latin (or Greek) as the main effects and with covariance adjustment for the total score (for parts 3-5) of the MLAT. The principal results are shown in Table 7.15. Data are given only for the Latin-No Latin main effect; although the Time of Beginning the major language produced highly significant effects, these are similar to those discussed earlier in this chapter and need not further concern us. The interaction effects between the two main effect variables were all insignificant.

The relevance of the covariance adjustment for language aptitude becomes apparent when the mean MLAT scores are examined for the groups with different experiences in Latin. It is evident that MLAT is correlated to a considerable extent with Latin study, probably significantly so. (This particular finding was not separately tested because it is not of direct concern to this investigation.) There is no sure explanation for this; the most probable is that students tend to self-select themselves for Latin study on the basis of their experiences with foreign languages, which are to some extent related to language aptitude. It is also possible that the study of Latin or ancient Greek may enhance some of the abilities intrinsic to language aptitude as measured by the MLAT.

In the table, the mean MLA skills test scores for the various groups are given without adjustment for language aptitude. At the right of the table, however, are given the "effects" of the Latin-No Latin variable after covariance adjustment for MLAT. These effects are the deviations of the adjusted mean scores, measured from the grand means of the relevant variable, associated with each sub-classification of the Latin-No Latin variable. The effects for any given sample and variable sum to zero. For example, for the French group, under the Listening column, the effects are -0.35, 0.91, 0.47, and -1.03. This means that the No Latin group's adjusted mean Listening score is .35 points under the adjusted grand mean, the High School Only group's adjusted mean score is .91 points above the adjusted grand mean, etc. That is, for this variable, having had no Latin depresses the mean score while having had Latin in high school raises it.

In the whole table, there is only one series of effects that are statistically significant (i.e. beyond chance fluctuations) at the 1% level, namely the effects for the Spanish Speaking score. Here, having had no Latin enhances the Speaking score, while having had it in high school or college depresses it. This finding would tend to support the hypothesis held by some foreign language teaching experts that exposure to the grammar-translation practices of most Latin teaching tends to transfer negatively to the audio-lingual language learning desired in modern language teaching. The hypothesis is based on the notion that grammar-translation methods incline the student to make too much reference to English (that is, to the native language) and thus inhibit the development of successful foreign-language study habits in contexts stressing the making of immediate foreign-language responses without reference to the native language. It is interesting that this significant effect occurs only in the Speaking test, which presumably demands fluent for-

Table 7.15

Effects Associated with the Study of Latin and/or Ancient Greek

Group (by time of study of Latin or Greek)	N	Mean MLA Skills Test Scores				MLAT Total	Effects after adjustment for MLAT			
		L	S	R	W		L	S	R	W
FRENCH MAJORS										
None	331	43.55	73.65	48.09	46.14	77.50	-0.35	-0.26	-1.26	-1.67
In high school only	611	44.33	73.05	48.98	47.28	80.37	0.91	-0.51	0.14	-0.36
In college only	41	45.20	76.73	51.22	50.37	80.85	0.47	0.77	0.50	0.85
In both h.s. & coll.	54	44.69	75.56	51.02	49.96	83.26	-1.03	0.00	0.62	1.18
	1037					F-ratio:	0.48	3.22*	1.86	3.04*
GERMAN MAJORS										
None	103	43.01	86.27	50.01	50.01	73.67	-0.74	-0.79	-2.09	-3.06
In high school only	168	42.89	86.36	50.64	52.02	76.98	-1.09	-0.65	-1.64	-1.35
In college only	21	45.81	88.24	56.57	58.76	88.14	1.83	1.44	3.73	4.41
	292					F-ratio:	0.96	0.23	3.17*	2.82
SPANISH MAJORS										
None	392	44.53	84.85	47.67	54.38	71.41	0.38	1.19	-0.22	-0.17
In high school only	304	44.43	82.45	47.83	54.47	76.65	0.49	-0.73	0.14	-0.29
In college only	64	44.72	85.30	49.16	56.62	73.22	-0.86	-0.46	0.08	0.46
	760					F-ratio:	0.28	4.73**	1.10	2.04

**p < .01; *p < .05

eign-language production; in the same group, the effect is reversed for the other tests (Listening, Reading, and Writing). The effect for the Writing test is nearly significant at the 5% level.

On the other hand, no such effect for the Speaking test occurs for the other languages. In French, which is cognate with Spanish, the Speaking test results show greatest positive effect of Latin study for those who studied it in college. The only agreement between the Speaking test results in French and in Spanish lies in the fact that in both cases there is a negative effect for the group that studied Latin in high school--again supporting the thesis mentioned above that early exposure to Latin may inhibit good modern foreign-language learning.

In all three languages, positive effects of Latin study are to be noted in the Writing test scores for those who studied Latin in college; the effect is statistically significant in French and nearly so in German and Spanish. Since the Writing test stresses grammar, this finding would support the idea that Latin study has at least some beneficial effect on those aspects of modern language study that are associated with grammatical analysis, thus reinforcing the arguments that teachers of classical languages have been advancing for many years.

On the whole, however, the effects of Latin study revealed in this analysis are rather slight--seldom approaching statistical significance. In contrast to variables such as Time of Beginning FL Study and Amount of Time Abroad, it is a very weak and unreliable influence. There is little point in trying to interpret MLA skills test scores with reference to the amount of Latin study the individual has had.

Chapter VIII

REGRESSION ANALYSES FOR PREDICTING MLA SKILLS TEST SCORES FROM COMPOSITES OF VARIABLES

1. Introduction

In Chapters VI and VII we have explored relationships between MLA skills test scores and a number of student and instructional variables that were thought to be possibly associated with them. Each such variable was taken up in turn. A drawback of that procedure was that the possible interrelationships among the associated background variables were generally not taken into account. If two such variables are strongly correlated, redundancy exists and one variable should not be considered without the other. The present chapter attempts to remedy this drawback by utilizing statistical techniques whereby all the interrelationships of relevant background variables are examined. The goal is to determine how well the MLA test scores can be "predicted" or "explained" by the background variables. To the extent that background variables are amenable to manipulation--e.g. by sending more students abroad, or by improving certain aspects of instruction, the results of these analyses may suggest ways of turning out better language students.

The techniques of analysis come under the heading of what is called linear multivariate analysis. Multiple regression and canonical correlation are the two techniques that are employed extensively in this chapter. We shall attempt to explain these techniques for the layman by carefully describing their application to some relatively simple tabulations developed in the present study. First, however, several statistical terms will be introduced and defined:

Predictor: a variable that is used as a basis for prediction. For example, we can use scores on the MLAT as a possible predictor of foreign language achievement as measured by one of the MLA skills tests. Use of the term "predictor" does not necessarily imply that the variable is actually efficient as a predictor; it merely implies that the variable is entered into a regression system (defined below) as a possible predictor.

Criterion (plural: criteria): a variable that one hopes to be able to predict. In the present study, the MLA skills test scores serve as the criteria, since we hope to "predict" their values for individual cases, and thus to "explain" their variation.

Correlation: a statistical measure of the association between two variables, e.g. between a predictor and a criterion. When the correlation coefficient is zero, it indicates no association between the variables in question. When it is unity (+1.00), it indicates that the two variables are perfectly correlated, that is, that every individual value of one of the variables is relatively just as high or low as the corresponding value on the other variable. When it is negative unity (-1.00), it indicates that the values are inversely correlated, that is, that each individual value of one of the variables is just as relatively high as the corresponding individual value of the other variable is relatively low, and vice versa. The simple correlation coefficient may take values anywhere between +1 and -1. Correlations are computed on the basis of a series of paired values. When a correlation is computed between a predictor and a criterion, it is sometimes called a validity coefficient. Often, one is interested in determining the statistical significance of a correlation, that is, in whether the actual correlation departs significantly from the range of values that one might obtain by chance, because any series of values paired at random will show some tendency to be correlated, positively or negatively. Generally one tests the significance of a correlation coefficient as departing from the value of .00 that would be postulated to occur, on the average, if there is not truly an association between the variables.

Regression: When two variables, X and Y, are imperfectly correlated, that is, with the absolute (non-signed) coefficient less than 1.00, it will be found that the values of

Y associated with a given value of X tend not to be as far, relatively, from the mean of Y as the value of X is from the mean of X for all cases. Thus, values of Y are said to regress toward the mean. The average degree to which this regression occurs can be depicted by a regression line joining the average or expected values of Y for given values of X. The slope of the regression line is a function of the degree of correlation. For the procedure to be used here, it is implied that the regression line is the straight line of best fit to the data.

Multiple regression: a generalization of simple regression to the case where Y is predicted from two or more variables, X_1, X_2, etc. The regression line becomes, geometrically, a plane (in the case of two predictor variables) or a hyperplane (for three or more predictor variables), and the slopes of the plane or hyperplane can be measured with respect to each one of the predictor variables.

Regression analysis: the process of computing the best possible prediction line, plane, or hyperplane, and its slopes with respect to the predictor variables.

Regression system: the results of regression analysis for a given set of data, including the correlations, regression coefficients, etc.

Multiple correlation: a coefficient, analogous to the simple correlation coefficient, that expresses the degree to which given data are approximated by the best regression line (plane, hyperplane) computed for the data. It measures, therefore, the extent to which a given criterion variable can be predicted from a series of predictor variables. The multiple correlation coefficient may range between 0 and 1.

Beta coefficient or Beta-weight: technically, a measure of the slope of the regression plane with respect to a given predictor variable (all variables being expressed in standardized form). It indicates the degree to which the given predictor variable uniquely contributes to the prediction of the criterion variable. Values are ordinarily between -1.00 and +1.00, like correlation coefficients, but may occasionally exceed these bounds.

Linear composite: If we have a series of variables, X_1, X_2, etc., we can add them up, with positive or negative weights (for multiplying constants) attached to them. The result is called a linear composite. For example, we might have reason to form the linear composite $Y' = (+.3)\ X_1 + (-.4)\ X_2$, where (+.3) and (-.4) are the weights. Then if for a given individual $X_1 = 40$ and $X_2 = 25$, the value of Y' for that individual is $(+.3)(40) + (-.4)(25) = 2$. Regression analysis can now be viewed as a process of determining the best weights to use in a linear composite for predicting Y from a series of variables X_1, X_2, etc. Multiple correlation is the correlation of the linear composite, so formed, with the criterion.

Canonical correlation: a special type of correlation expressing the optimal relation between two sets of variables, Y_1, Y_2, etc., on the one hand, and X_1, X_2, etc., on the other by forming linear composites for each set and determining the optimal weights to use in each composite. It is the correlation between the linear composites so formed.

We shall now turn to the analysis of certain data from the present study that we will use as an illustration of multiple regression analysis and canonical correlation. We first analyze the data by multiple regression. The canonical regression is introduced in section 4.

2. Multiple Regression Prediction of MLA Skills Test Scores from MLAT and Three Background Variables, French Majors

The complete regression analysis data for 1039 "regular" cases in French whose data on all the relevant variables were complete is shown in Table 8.1, which we will explain in detail.

Table 8.1

Multiple Regression Analysis for Predicting MLA Skills Test Scores from MLAT Scores and Three Background Variables

French Majors, "Regular" Cases (N = 1039)

	Criteria: MLA Skills Test							
	Listening		Speaking		Reading		Writing	
Predictors	r	β	r	β	r	β	r	β
1 MLAT-3 (Spelling Clues)	.23**	.17**	.20**	.17**	.27**	.19**	.26**	.17**
2 MLAT-4 (Words in Sentences)	.15**	.09**	.09**	.06*	.26**	.19**	.28**	.23**
3 MLAT-5 (Paired-Associates)	.13**	.05*	.05	-.01	.18**	.08**	.17**	.06*
4 Time Began (see coding below)	-.22**	-.18**	-.22**	-.18**	-.15**	-.13**	-.23**	-.20**
5 Time Abroad (" " ")	.50**	.47**	.39**	.37**	.34**	.32**	.36**	.34**
6 Use of FL at Home (" ")	.14**	.10**	.20**	.16**	.11**	.10**	.13**	.11**
Multiple Correlation	.59**		.51**		.51**		.55**	

b-Weights

	Listening	Speaking	Reading	Writing
1	.13	.16	.17	.16
2	.11	.09	.25	.35
3	.08	-.02	.14	.11
4	-2.49	-3.08	-2.04	-3.60
5	4.51	4.30	3.40	4.01
6	2.16	4.19	2.32	2.97
Intercept	29.95	60.57	28.96	26.10

Intercorrelations of Predictors

	1	2	3	4	5	6
1	1.00	.28**	.22**	-.07*	.01	-.01
2	.28**	1.00	.32**	.08**	.02	-.07*
3	.22**	.32**	1.00	.03	.03	-.05
4	-.07*	.08**	.03	1.00	-.06*	-.09**
5	.01	.02	.03	-.06*	1.00	.06*
6	-.01	-.07*	-.05	-.09**	.06*	1.00
Mean	27.59	32.33	19.71	2.03	1.74	1.12
S.D.	9.81	6.26	4.90	0.54	0.80	0.36

Intercorrelations of Criteria

	1	2	3	4
1	1.00	.69**	.73**	.76**
2	.69**	1.00	.58**	.65**
3	.73**	.58**	1.00	.80**
4	.76**	.65**	.80**	1.00
Mean	44.14	73.52	48.89	47.18
S.D.	7.59	9.27	8.51	9.53

**$p < .01$, *$p < .05$ for hypothesis value of zero.

Coding

Time Began	Time Abroad	Use of FL at Home
1 Started in grade school	1 Never Abroad	1 Never
2 Started in high school	2 Summer or Tour Abroad	2 Occasionally
3 Started in college	3 Year of Study Abroad	3 Frequently

The table concerns the prediction of MLA Skills test scores, one by one, from 6 predictor variables that we know from previous chapters are relevant to foreign language attainment. In a sense, it is an assemblage of four separate tables, one for each of the MLA skills tests, because each of these is to be predicted separately from the same predictors. The basic data from which the analysis starts are the intercorrelations of the predictors (shown at the bottom of the table) and the correlations of the predictors with each of the criteria (shown in the columns labeled "r" at the top of the table). In the present case there are only a few small significant correlations among the predictors, chiefly among the MLAT tests; nevertheless, each of the predictors shows a significant correlation with one or more of the criteria. The large number of cases makes a correlation of only .06 significant at the 5% level, and .08 significant at the 1% level. (That is to say, a correlation as large as .06 in absolute magnitude would occur by chance only 5% of the time, and a correlation as large as +.08 would occur by chance only 1% of the time.)

The actual computations, of course, are complex and are not shown here. Their purpose is to determine weights (b-weights) which when multiplied by individual scores will yield a linear composite that will best predict a given criterion. This is done through an intermediate stage that determines the "beta-weights" that would be applied to "standardized variables" for the same purpose. A standardized variable is a variable where the score is expressed in terms of number of standard deviation units from the mean. Thus, if a variable has a mean of, say, 40, and a standard deviation of 5, a score of 50 on that variable would be expressed as +2.00 in standardized form, because it is 2 standard deviation units above the mean. Similarly, a score of 25 on the variable would be expressed as -3.00 in standardized form. Beta-weights, then, are the weights for predictor variables expressed in standardized form. They constitute information as to the relative extent to which a given predictor variable contributes uniquely to the prediction of the criterion. They take account of the correlations among the predictor variables; thus, if two highly correlated predictor variables correlate about equally with the criterion, generally only one of these will be assigned a high beta-weight, since the other variable does not contribute much unique information to the prediction after the first variable is used.

Let us look, now, at the correlations and beta-weights (β) for the prediction of the French Listening test score. As noted earlier, each of the predictors has a significant correlation with this criterion; Time Abroad has the highest correlation, .50. Time Began has a negative correlation, meaning that high values of that predictor are associated with low values of the criterion, and vice versa. It will be recalled that Time Began is coded in such a way that "1" represents starting study of the FL in grade school, "2" represents starting in high school, and "3" represents starting in college. Chapter 7 showed that in general those who start in college have lower skills test scores than those who start in high school, and these in turn have lower scores than those who start in grade school. The negative correlation found here confirms these results. The remaining predictors have positive correlations with the criterion.

In the present case, the beta-weights correspond roughly to the correlations because of the small intercorrelations among the predictors. The beta-weights for MLAT-4 and MLAT-5 are decreased relative to that of MLAT-3, however, because of the intercorrelations of the MLAT subtests, and MLAT-3 gets assigned the higher beta-weight of the three. From a relative standpoint, it now appears that Time Abroad contributes most to the prediction, MLAT-3 next most, Time Began third most, and Use of FL at Home fourth most. That is, these numbers indicate the extent to which variation in Listening scores is "explained" by the predictor variables.

Similar interpretations can be derived from the figures for the prediction of Speaking, Reading, and Writing scores. Time Abroad is universally the best predictor of the set, yet each of the other predictors generally makes some significant independent contribution. It is noteworthy that MLAT-4, a test of "grammatical sensitivity", makes a greater contribution to Reading and Writing scores than it does to Listening and Speaking scores.

The b-weights given below the correlations and beta-weights represent the beta-weights translated into values that are directly applicable to raw scores. For example, suppose we had an individual whose scores on the predictors were as given below:

Predictor	Individual A's score	Listening b-weight	Score x b-weight
MLAT-3	35	.13	4.55
MLAT-4	27	.11	2.97
MLAT-5	21	.08	1.68
Time began	2	-2.49	-4.98
Time abroad	3	4.51	13.53
Use of FL at home	1	2.16	2.16
		Sum	19.91
		Intercept	29.95
	Predicted Listening Score		49.85

Then we could, as the table shows, multiply each score by the corresponding b-weight, sum the results, and add the "intercept" to the result to obtain the best prediction of the Listening score. This predicted score is technically a linear composite of the predictors. The prediction is "best" in the sense that there is the smallest standard deviation of errors made in prediction, an "error" being the algebraic difference between the actual score and the predicted score. For the present data, it is only an academic exercise to compute predicted scores and in the future we will drop the reporting of b-weights, since we are mainly interested in determining the relative extents to which criteria are explained by the predictors, as indicated by beta-weights.

We will, however, continue to report multiple correlations (R), since these indicate the overall extent to which a criterion variable can be predicted by the predictor variables through multiple regression analysis. For the Listening criterion, Table 8.1 shows the multiple correlation to be .59, and it is very significantly different from zero (as shown by the double asterisk). Technically, this multiple correlation is the simple correlation between the criterion variable and the linear composite formed by weighting the predictor variables by the b-weights (or by the beta-weights if the predictors are expressed in standardized form). That is, if we were to form a linear composite as shown above for each of the 1039 individuals in the sample, there would be a correlation of .59 between these predicted Listening scores and the actual Listening scores.

By adding more predictors, we will hope to predict the Listening scores even more accurately, that is, explain a greater proportion of their variance. This in fact will be done later in this chapter.

Table 8.1 includes also the intercorrelations of the criteria. These intercorrelations are not actually used in the computation of the multiple regressions since each of the criteria is dealt with separately. However, the relatively high intercorrelations of the criteria explain why the multiple regression systems for them are fairly similar.

The only term in Table 8.1 that has not been explained is "intercept". We will only say that it is a number that must be included in the linear composite to make the mean predicted score equal to the mean of the actual score.

3. Multiple Regression Techniques Applied to Prediction of Each MLA Skills Test Variable, in Each Language

An explanation of the multiple regression technique having been given, we can proceed to consider the results of its application to various sets of data obtained for "regular" cases in the several languages. Because the stability of the results over languages is of more interest, possibly, than the comparative results for the different skills, we will present results for each MLA skills test in turn.

Table 8.2, having to do with the prediction of Listening test scores, contains in its first column beta-weights and the multiple correlation for Listening test scores already presented in Table 8.1 for the French sample. It also contains, however, analogous results for other language groups and for different sets of variables. There are five main sets of results:

(1) Data for complete samples in French, German, Russian, and Spanish, using three MLAT subtests and three variables that summarize the backgrounds of the students. (The results for French were presented in the preceding section.)

(2) Data for complete samples in French, German, Russian, and Spanish, using MLAT total score, amount of time abroad, the extent the FL is used at home, and 5 variables that present in some detail the training histories of the student.

(3) Data for French and Spanish samples that started study of their major foreign language in elementary school ("grade school," as it was identified in the student questionnaire). (There were insufficient numbers of cases to make comparable analyses in German or Russian.)

(4) Data for French, German, and Spanish samples that started study of their major foreign language in high school. (Insufficient data were available for an analysis for the Russian group.)

(5) Data for samples in each of the four languages that started study of their major foreign language in college.

For the last three sets of data, the intent was to select all applicable variables that promised to be relevant and that were relatively independent of each other. However, because of the small numbers of cases of "grade school starters," certain training history variables were omitted for those cases. Number of semesters of grade school was omitted for the high school starters because it was irrelevant, and data on high school training history were omitted for the college starters for a similar reason.

For clarity of presentation and for economy of space, simple correlation coefficients are not given.

A word must be said about statistical significance. When the number of cases is relatively small, as it is for the Russian majors, statistics such as correlations and beta-weights must be larger in absolute magnitude to be significantly different from zero, since random fluctuations could produce fairly large values even when the "true" population values are zero. In the data for the Russian majors, a number of beta-weights as high as .24 fail to be significantly different from zero at the 5% level of significance. Nevertheless, these values are the best estimates we have of the values we might obtain if we had a much larger sample, and can be respected as indicating trends of possible interest. The number of predictor variables also affects statistical significance; each additional variable has the same effect as dropping one case.

Even from the 6 variables used in the first set of data, the MLA Listening test scores are substantially well predictable. The MLAT subtests make significant independent contributions in many instances. The Time Began variable is a significant predictor for French and Spanish, and the beta-weight nearly reaches significance in Russian. The amount of time abroad is clearly the best of the predictors in each case, and Use of the FL at Home generally runs second.

Prediction is even better when more variables are added to present the training history in more detail, substituting for the gross Time Began variable in the first set of data. Again, Time Abroad is the strongest predictor, and Use of the FL at Home continues to be influential in the prediction. The aggregate effects of the variables representing the training history are not large, however, being overridden by Time Abroad and Use of FL at Home. It would seem that the amount of time the student spends on language study,

Table 8.2

Multiple Regression Analyses for the MLA Listening Test (Beta Weights and Multiple R's)

Predictors \ Language	Complete Samples, MLAT & Background				Complete Samples, MLAT, Background, & History				Grade School Starters		High School Starters			College Starters			
	F	G	R	S	F	G	R	S	F	S	F	G	S	F	G	R	S
MLAT-3	.17**	.02	-.09	.15**	--	--	--	--	.20**	.04	.14**	.08	.15**	.13	-.04	-.06	.08
MLAT-4	.09**	.11*	.22	.10**	--	--	--	--	-.12	.15	.13**	.04	.15**	.07	.21**	.27	-.03
MLAT-5	.05*	-.01	-.13	-.01	--	--	--	--	.06	.04	.07*	.00	.00	.02	-.01	-.14	-.08
MLAT-Total	--	--	--	--	.23**	.09	.02	.17**	--	--	--	--	--	--	--	--	--
Time Began	-.18**	-.01	-.16	-.10**	--	--	--	--	--	--	--	--	--	--	--	--	--
Time Abroad	.47**	.60**	.24*	.40**	.43**	.53**	.30*	.31**	.48**	.46**	.42**	.46**	.31**	.35**	.56**	.25	.31**
Use FL at Home	.10**	.20**	.23*	.27**	.08**	.18**	.19	.27**	.11	.15	.07*	.25**	.27**	.08	.11	.26	.18*
Sem. G. S.	--	--	--	--	.07**	.10*	.05	.08*	.09	.11	--	--	--	--	--	--	--
Sem. H. S.	--	--	--	--	.12**	-.01	.13	.07*	-.05	-.09	.08*	.00	.10*	--	--	--	--
Teacher's Lang. (H. S.)	--	--	--	--	--	--	--	--	.12	.13	.03	.04	.00	--	--	--	--
Hrs. Coll. Lang.	--	--	--	--	.00	.13**	.09	-.02	-.06	-.04	.01	.15	-.01	.07	.09	.09	-.07
Teacher's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	.09	-.18	.04	-.03	.07	.10	-.10	.06	.25**
Student's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	--	--	.05	.03	.04	-.07	.17	.11	.01
Hrs. Coll. Lit.	--	--	--	--	.21**	.11*	.18	.20**	.06	.08	.19**	.19*	.19**	.13	-.01	.16	.17*
Teacher's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	.07	.22*	.05	.12	-.05	.34**	.19*	.16	.01
Student's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	--	--	.08	-.05	.12	-.13	.01	.03	.03
Hrs. Misc. Courses	--	--	--	--	.01	.08	-.12	.09**	--	--	--	--	--	--	--	--	--
Multiple R	.59**	.66**	.44*	.55**	.63**	.68**	.43*	.59**	.64**	.68**	.66**	.67**	.61**	.59**	.73**	.58**	.57**
N	1039	289	80	764	1039	289	80	763	139	91	733	131	518	167	151	58	154

$^{**}p < .01$; $^{*}p < .05$

Table 8.3

Multiple Regression Analyses for the MLA Speaking Test (Beta Weights and Multiple R's)

Predictors \ Language	Complete Samples, MLAT & Background				Complete Samples, MLAT, Background, & History				Grade School Starters		High School Starters			College Starters			
	F	G	R	S	F	G	R	S	F	S	F	G	S	F	G	R	S
MLAT-3	.17**	.07	-.16	.13**	--	--	--	--	.25**	.13	.13**	.07	.13**	.16*	.09	-.09	.04
MLAT-4	.06*	.07	.30*	.07*	--	--	--	--	-.18*	.03	.08*	.05	.12**	.13	.12	.35*	.00
MLAT-5	-.01	-.12*	-.14	-.03	--	--	--	--	-.03	-.09	.00	-.14	.01	-.09	-.11	-.25	-.09
MLAT-Total	--	--	--	--	.17**	.03	.00	.14**	--	--	--	--	--	--	--	--	--
Time Began	-.18**	-.07	.01	-.06*	--	--	--	--	--	--	--	--	--	--	--	--	--
Time Abroad	.37**	.49**	.27*	.30**	.31**	.47**	.33**	.25**	.34**	.36**	.33**	.40**	.25**	.27**	.43**	.22	.25**
Use of FL at Home	.16**	.25**	.28*	.33**	.14**	.24**	.22	.32**	.15	.33**	.17**	.28**	.30**	.08	.16*	.27	.17*
Sem. G. S.	--	--	--	--	.09**	.08	-.03	.11**	.08	.11	--	--	--	--	--	--	--
Sem. H. S.	--	--	--	--	.15**	.02	.06	.02	-.08	-.19	.10**	-.09	.06	--	--	--	--
Teacher's Lang. (H. S.)	--	--	--	--	--	--	--	--	.23*	.05	.00	.07	.04	--	--	--	--
Hrs. Coll. Lang.	--	--	--	--	.05	.04	.06	.00	.16*	.08	.01	.13	-.04	.13	-.03	.08	.08
Teacher's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	-.03	-.15	.02	-.23	.05	.12	-.15	-.17	.09
Student's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	--	--	.12	.17	.00	-.12	.17	.22	.10
Hrs. Coll. Lit.	--	--	--	--	.12**	.07	-.12	.08*	-.12	-.07	.11**	.16	.09*	.06	.00	-.15	.00
Teacher's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	.08	.18*	.03	.04	-.05	.15	.19	.16	.06
Student's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	--	--	-.01	.03	.07	.12	-.03	-.16	.08
Hrs. Misc. Courses	--	--	--	--	.10**	.01	-.16	.03	--	--	--	--	--	--	--	--	--
Multiple R	.51**	.60**	.50**	.49**	.52**	.59**	.48*	.50**	.58**	.73**	.52**	.66**	.48**	.56**	.58**	.59**	.47**
N	1039	289	80	764	1039	289	80	763	139	91	733	131	518	167	151	58	154

$^{**}p < .01$; $^{*}p < .05$

in terms of semesters or course hours from grade school on, is far less important than the kind of exposure to the foreign language that he can get at home or through study abroad. Language aptitude plays a small but generally significant role.

In the case of the grade school starters, where amount of school exposure has been largely evened out, Time Abroad is practically the only variable that makes any significant difference in Listening score. Nevertheless, the overall prediction remains highly accurate. It is of interest that in the case of the Spanish grade school starters, the extent to which the instructors in college literature courses use the foreign language in class is a significant predictor.

Multiple correlations between .61 and .67 are found for the prediction of Listening scores of students who started in high school. Language aptitude plays a small but significant role. Besides the major predictors (Time Abroad and Use of the FL at Home), Hours of College Literature is seen to be an important variable, consistently over the language groups.

The pattern of results is similar for those who started their foreign language in college, except that the role of the teacher's use of the language in the classroom becomes even more significant. The multiple correlations remain high even when the group becomes relatively homogeneous in training history.

Table 8.3 presents results for the MLA Speaking test. The overall accuracy of prediction, as reflected in the multiple correlations, is generally slightly smaller than in the case of the Listening test; this may be partly a function of lower reliability of the Speaking test criterion. Otherwise, the pattern of results is highly similar to that for the Listening test.

The results in Table 8.4, for the MLA Reading test, are similar in most respects to those for the Listening and Speaking tests, except that language aptitude variables, especially MLAT-4 (a test of "grammatical sensitivity")have more role than before. Also, the beta-weights for Hours of College Literature tend to increase slightly, as we might expect them to in view of the fact that college literature courses are designed to encourage reading skill. Time Abroad, for some reason, fails to attain significance for the college starters, possibly because individuals starting in college are less likely to have had experience abroad (as was shown in Chapter VII).

Prediction of scores on the Writing test (Table 8.5) depends on much the same variables as was found to be the case for the Reading test, with certain language aptitude variables slightly more emphasized.

Somewhat surprisingly, there is only a scattering of instances in which the beta-weight for Hours of College Language courses (as opposed to literature courses) attains significance. If anything, the trend is for these values to be negative, especially for students starting FL study early. This does not signify necessarily that college language courses are of no value. It may simply signify that individuals who arrive at college relatively less well prepared in language skills generally have to take more such courses in order to achieve the competence prerequisite for literature courses, and these are in general the poorer students.

The results shown in Tables 8.2 through 8.5 indicate, in summary, that variations in the amounts of skill that college language majors attain in their major foreign language can be rather well accounted for by variations in their language aptitude and the amounts of exposure they have had to the language either in or out of school.

Table 8.4

Multiple Regression Analyses for the MLA Reading Test (Beta Weights and Multiple R's)

Predictors \ Language	Complete Samples, MLAT & Background				Complete Samples, MLAT, Background, & History				Grade School Starters		High School Starters			College Starters			
	F	G	R	S	F	G	R	S	F	S	F	G	S	F	G	R	S
MLAT-3	.19**	.04	-.11	.16**	--	--	--	--	.22**	.05	.17**	.03	.17**	.13	.02	-.04	.05
MLAT-4	.19**	.24**	.29*	.15**	--	--	--	--	-.03	-.01	.21**	.19*	.22**	.22**	.34**	.39*	.06
MLAT-5	.08**	-.08	-.25*	.02	--	--	--	--	.09	.09	.13**	-.07	.02	-.05	-.08	-.24	.01
MLAT-Total	--	--	--	--	.33**	.16**	-.03	.24**	--	--	--	--	--	--	--	--	--
Time Began	-.13**	-.06	-.24*	-.05	--	--	--	--	--	--	--	--	--	--	--	--	--
Time Abroad	.32**	.52**	.14	.34**	.29**	.48**	.25*	.27**	.41**	.38**	.27**	.36**	.29**	.15	.53**	.15	.22**
Use of FL at Home	.10**	.17**	.20	.30**	.06*	.14**	.12	.28**	.10	.11	.06	.23**	.27**	.10	.07	.23	.20**
Sem. G. S.	--	--	--	--	.07**	.09	.06	.06*	.14	.03	--	--	--	--	--	--	--
Sem. H. S.	--	--	--	--	.04	.05	.15	-.01	-.04	-.08	.08*	.03	.05	--	--	--	--
Teacher's Lang. (H. S.)	--	--	--	--	--	--	--	--	.03	.08	-.05	.07	.00	--	--	--	--
Hrs. Coll. Lang.	--	--	--	--	-.05	.09	-.06	-.10**	-.08	-.08	-.06	.17	-.09*	-.01	.01	-.13	-.18*
Teacher's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	.06	-.28**	.09	-.06	.03	.04	-.07	.01	.26*
Student's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	--	--	.01	.08	.05	-.05	.10	.13	-.07
Hrs. Coll. Lit.	--	--	--	--	.23**	.13*	.06	.20**	.15	.14	.22**	.28**	.16**	.18*	.01	.10	.21**
Teacher's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	-.01	.19	.03	-.07	-.03	.22	.14	-.24	.15
Student's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	--	--	.00	.04	.09	.00	-.03	.45	-.03
Hrs. Misc. Courses	--	--	--	--	-.02	.07	-.11	.07*	--	--	--	--	--	--	--	--	--
Multiple R	.51**	.60**	.48**	.53**	.56**	.60**	.35 (p = .27)	.57**	.58**	.61**	.60**	.61**	.60**	.48**	.67**	.58**	.55**
N	1039	289	80	764	1039	289	80	763	139	91	733	131	518	167	151	58	154

**p < .01; *p < .05

Table 8.5

Multiple Regression Analyses for the MLA Writing Test (Beta Weights and Multiple R's)

	Complete Samples, MLAT & Background				Complete Samples, MLAT, Background, & History				Grade School Starters		High School Starters			College Starters			
Predictors \ Language	F	G	R	S	F	G	R	S	F	S	F	G	S	F	G	R	S
MLAT-3	.17**	-.02	-.18	.21**	--	--	--	--	.10	.07	.16**	.00	.20**	.12	-.05	-.11	.19*
MLAT-4	,23**	.33**	.24	.18**	--	--	--	--	.13	.11	.24**	.32**	.23**	.22**	.41**	.15	.05
MLAT-5	.06*	-.09	-.21	.01	--	--	--	--	.03	.13	.10**	-.13	.03	-.03	-.05	-.16	-.09
MLAT-Total	--	--	--	--	.33**	.17**	-.13	.29**	--	--	--	--	--	--	--	--	--
Time Began	-.20**	-.03	-.14	-.06	--	--	--	--	--	--	--	--	--	--	--	--	--
Time Abroad	.34**	.50**	.13	.32**	.30**	.46**	.18	.24**	.42**	.31**	.28**	.40**	.26**	.18*	.41**	.14	.21**
Use of FL at Home	.11**	.18**	.15	.23**	.08**	.14**	.10	.23**	.12	.10	.08**	.19*	.22**	.10	.10	.13	.13
Sem. G. S.	--	--	--	--	.09**	.09	.09	.06*	.09	.10	--	--	--	--	--	--	--
Sem. H. S.	--	--	--	--	.10**	.05	.21	.06	.04	.15	.06	.01	.08*	--	--	--	--
Teacher's Lang. (H. S.)	--	--	--	--	--	--	--	--	.05	-.17	-.01	.06	.00	--	--	--	--
Hrs. Coll. Lang.	--	--	--	--	-.03	.18**	.22	-.03	.04	-.07	-.04	.11	-.10*	.02	.11	.17	.06
Teacher's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	.03	-.21	.14*	-.04	.03	.00	.00	.02	.24*
Student's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	--	--	-.02	.10	.11	-.02	.07	-.02	-.03
Hrs. Coll. Lit.	--	--	--	--	.20**	.12*	.12	.20**	-.02	.03	.21**	.15	.18**	.17*	.04	.03	.09
Teacher's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	.09	.21*	.06	.09	.01	.27*	.19	.26	.09
Student's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	--	--	-.02	-.05	.06	-.01	.03	.16	.06
Hrs. Misc. Courses	--	--	--	--	.01	.08	-.14	.07*	--	--	--	--	--	--	--	--	--
Multiple R	.55**	.61**	.42*	.52**	.58**	.60**	.42 (p = .08)	.56**	.57**	.53**	.60**	.64**	.63**	.53**	.68**	.59**	.53**
N	1039	289	80	764	1039	289	80	763	139	91	733	131	518	167	151	58	154

**p < .01; *p < .05

4. Canonical Regression Analysis of MLA Skills Test Scores with MLAT and Three Background Variables as Predictors, French Majors

In this section we introduce a somewhat more elaborate technique of statistical analysis, canonical regression analysis, and apply it to the same data that were discussed in section 2. Certain purposes are attainable with canonical regression analysis that are not attainable with ordinary multiple regression analysis. It is applicable only when there are two or more criterion variables (that are not perfectly correlated). One purpose it serves is to yield information as to what extent it is possible to predict some combination (linear composite) of criterion variables better than any one of them severally can be predicted. Further, it allows us to explore the possibility that these criteria can be meaningfully predicted in different ways. For example, if there are two criterion variables, it may be that some predictors predict the sum of these variables (or some sum with positive weights), and other predictors predict the algebraic difference between the variables (or more generally, a linear composite of the criteria such that one variable has a positive weight and the other variable has a negative weight).

Technically, a canonical correlation is a simple correlation between two linear composites, one constituted from a set of predictors, and one from a set of criteria. The weights for these composites are computed in such a way that they maximize the correlation between the composites. However, it is often the case that two or more sets of linear composites can be formed in such a way as to maximize their intercorrelations. Such sets of composites are formed in a series: the first canonical correlation, the second canonical correlation, etc., each successive correlation being formed from the residual covariance, that is, the covariation that remains in the data after the effects of the previous canonical correlations have been removed.

The weights for the linear composites are computed only in "normalized" form, i.e., the sum of their squares are equal to unity. Otherwise, they are analogous to beta-weights, except that they are computed in two sets for any given canonical correlation--one for the predictor variables and one for the criterion variables. The statistical significance of the canonical correlations can be tested, but of course the same constraints concerning the number of cases and the number of variables apply. As the number of cases decreases and the number of variables increases, the canonical correlation must be ever larger in absolute magnitude to attain statistical significance. No statistical tests of canonical regression weights were available in the computer programs we used.

We may now examine Table 8.6, which presents a canonical regression analysis for the data of Table 8.1. The intercorrelations of the criterion variables as well as those of the predictor variables are used in canonical regression analysis, but to save space they are not repeated in Table 8.6. It was found that four canonical correlations were significant, although the size of the last two is very small. It is a characteristic of statistical data that very small effects can be significant when there are very large numbers of cases, as in the present instance.

The first canonical correlation system, with a coefficient of .61, obviously refers to the overall predictability of the criterion tests, with positive weights, from the predictor variables weighted as indicated. The regression coefficients for the predictors (listed at the left of the table) are in the column headed P, while those for the criteria (as listed at the right of the table) are in the column headed C. The Listening test turns out to be the most predictable of the criteria; this finding agrees with that of the multiple regression analysis. Thus, the best overall measure of French language competence, from these results, would be one in which standardized scores on Listening, Speaking, Reading, and Writing are added up with weights of .87, .33, .03, and .37 respectively.

The second canonical can be interpreted as representing a tendency for the superiority of a student in Reading and Writing (positive weights) over his ability in Listening and Speaking (negative weights) to be predicted chiefly by MLAT-4 (Words in Sentences), and secondarily by the amount of time he has not spent abroad (since Time Abroad has a

Table 8.6

Canonical Regression Analysis for Data of Table 8.1
(MLAT Subtests and Three Background Variables as
Predictors, Four MLA Skills Tests as Criteria)
French Majors (N = 1039)

	Successive Canonical Correlations and Regression Weights								
	I		II		III		IV		
	P	C	P	C	P	C	P	C	Criteria
1 MLAT-3 (Spelling Clues)	.33	.87	.10	-.59	.05	-.60	.68	-.30	Listening
2 MLAT-4 (Words in Sentences)	.23	.33	.85	-.35	.16	.59	-.31	.30	Speaking
3 MLAT-5 (Paired-Associates)	.08	.03	.23	.37	-.48	-.34	.05	.69	Reading
4 Time Began	-.35	.37	-.07	.63	-.50	.42	.53	-.59	Writing
5 Time Abroad	.81		-.46		-.40		-.11		
6 Use of FL at Home	.23		-.05		.58		.39		
Canonical R	.61**		.30**		.15**		.11**		

**p < .01. Significance reported only for canonical correlations.

negative weight on the predictor side). This finding has much logical appeal: MLAT-4 measures "grammatical sensitivity," which might be expected to enhance a student's performance in Reading and Writing, particularly the latter, calling as it does on high competence in the grammatical aspects of the language; and students who have not been abroad might be expected to devote themselves more to the written aspects of a language or at least to have achieved relatively less competence in spoken aspects.

The third canonical, with a relatively small coefficient of .15, appears to refer to a prediction of the superiority of active performances in the language (Speaking and Writing having positive weights) over passive performances (represented by Listening and Reading). This kind of superiority is associated with those who began French study early rather than late, who have homes where the foreign language is used occasionally or even frequently, who are poor on rote memory, but who have not spent much time abroad. The negative weight of time abroad is a trifle puzzling in view of the pattern of other coefficients, for one would expect students who have spent much time abroad to be superior in active skills. But then, this canonical applies only after the effects of the first two have been extracted, and Time Abroad was already shown to be associated with Listening and Speaking competence.

The fourth canonical, though significant, is difficult to interpret. It seems to show that after all the other covariation is taken account of, there is a tendency for students with high scores on MLAT-3 and who began French late show superiority in Reading and (to some extent) Speaking, as contrasted with Listening and Writing.

5. Canonical Regression Techniques Applied to Prediction of Linear Composites of MLA Skills Test Scores

Canonical regression techniques were applied to the same samples of students, with the same sets of variables, as the multiple regression techniques. Each canonical regression system contained four canonical correlations, varying numbers of which were statistically significant, depending largely on the numbers of cases involved. The character of a linear composite determined by a canonical correlation is not necessarily associated with the order (first, second, third, or fourth) in which it is extracted from a given set of data. Therefore it seemed wise to examine all the data in terms of the patterns of coefficients for the criterion variables, sorting the canonical composites on the criterion side into roughly similar patterns. With allowances made for sampling fluctuations arising from limited numbers of cases in some instances, it was possible to establish four such patterns, each named in terms of the combination of MLA skills tests it represented:

(1) A general foreign language competence factor. These were criterion composites for which all or nearly all the coefficients were positive (or at least unidirectional); in most cases two or three of the coefficients were fairly substantial in magnitude. In almost all instances these composites were those yielded by the first canonical correlation. Sixteen of the 17 canonical correlations were statistically significant at the 5% level or better. These composites were therefore interpreted as representing a general competence factor in the relevant language, inasmuch as they set forth the way in which the criterion variables could be optimally weighted to be predicted by a set of variables known to be univocal in predicting the several measures of competence. This general foreign language competence factor corresponds to the general factor found in the intercorrelations of the MLA skills test scores in Chapter IV. The canonical regression coefficients for the several sets of data analyzed are to be found in the lower four rows of Table 8.7. It will be seen that the coefficients for the Listening test are generally the highest, agreeing with previous findings that the Listening test is the best overall measure of skill in the foreign language.

(2) A factor representing superiority in written over spoken skills. This will be termed the WR factor, denoting the fact that the coefficients are usually positive for the Writing and Reading tests and negative for the other two tests. These coefficients are shown in the last four rows of Table 8.8. The directionality chosen was largely arbitrary: the factor could equally well have been termed the LS (Listening and Speaking) factor, with the coefficients reversed in sign, except that it seemed desirable to retain a positive direction for certain important predictors. (The directionality of a linear composite is arbitrary, except that when the criterion coefficients are reversed in sign, the coefficients for the predictors must also be reflected. Reflections of sign were made for both predictors and criteria so that as much agreement as possible was obtained among the canonical composites obtained for the several sets of data.) This factor may be regarded as a measure of the extent to which a student's skills in Reading and Writing, on the one hand, are different from his skills in Listening and Speaking. With the directionality chosen for the coefficients, it is a measure of the student's superiority in written over spoken skills. In most instances, this factor was associated with the second canonical correlation, indicating that it was the second most important source of variance in the data. Twelve of the 17 canonical correlations were statistically significant at the 5% level or better.

(3) A factor representing superiority in productive skills. This will be termed the SW factor, reflecting its largely positive coefficients for Speaking and Writing as contrasted to largely negative coefficients for Listening and Reading. The coefficients are shown in the last four rows of Table 8.9. As in the case of the second factor, the directionality chosen was arbitrary. With signs reversed, it could equally well be interpreted as a factor representing superiority in receptive skills. It was associated about equally with the third and fourth canonical correlations, and only five out of 17 of these canonical correlations were significant at the 5% level or better. Thus, this dimension is a relatively weak source of variance. But it nevertheless makes some psychological sense.

(4) A factor representing superiority in reading and speaking. This is here termed the RS factor. The coefficients for the criterion variables are shown in the last four rows of Table 8.10. This factor was a weak source of variance in the data, as all the canonical correlations were relatively small in magnitude, and only four out of 17 of

Table 8.7

Canonical Regression Results for the General FL Competence Factor

Variable		Complete Samples, MLAT & Background				Complete Samples, MLAT, Background, & History				Grade School Starters		High School Starters			College Starters			
	Canonical:	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	II	I
	Language:	F	G	R	S	F	G	R	S	F	S	F	G	S	F	G	R	S
Predictors																		
MLAT-3		.33	.06	-.26	.30	--	--	--	--	.38	.16	.30	.14	.33	.29	-.05	-.17	.05
MLAT-4		.23	.15	.57	.23	--	--	--	--	-.19	.06	.33	.04	.39	.18	.39	.49	.01
MLAT-5		.08	-.07	-.34	-.01	--	--	--	--	.08	-.07	.17	-.09	.04	-.05	-.04	-.35	-.04
MLAT-Total		--	--	--	--	.50	.12	.11	.41	--	--	--	--	--	--	--	--	--
Time Began		-.35	-.05	-.17	-.14	--	--	--	--	--	--	--	--	--	--	--	--	--
Time Abroad		.81	.92	.46	.70	.72	.89	.75	.57	.80	.66	.73	.73	.54	.64	.82	.38	.55
Use of FL at Home		.23	.35	.50	.59	.16	.32	.53	.56	.22	.42	.16	.44	.49	.16	.15	.41	.40
Sem. G. S.		--	--	--	--	.15	.16	-.11	.16	.17	.15	--	--	--	--	--	--	--
Sem. H. S.		--	--	--	--	.21	-.01	.04	.07	-.09	-.33	.15	-.08	.15	--	--	--	--
Teacher's Lang. (H. S.)		--	--	--	--	--	--	--	--	.23	.21	.01	.08	.00	--	--	--	--
Hrs. Coll. Lang.		--	--	--	--	-.01	.18	.12	-.09	-.01	.05	-.02	.22	-.14	.20	.13	.20	-.32
Teacher's Lang. (Coll. Lang.)		--	--	--	--	--	--	--	--	.10	-.30	.13	-.20	.08	.22	-.12	.08	.49
Student's Lang. (Coll. Lang.)		--	--	--	--	--	--	--	--	--	--	.06	.15	.13	-.18	.22	.10	-.05
Hrs. Coll. Lit.		--	--	--	--	.38	.16	-.16	.37	.05	.02	.40	.26	.34	.19	.00	.19	.41
Teacher's Lang. (Coll. Lit.)		--	--	--	--	--	--	--	--	.12	.30	.10	.19	-.04	.53	.27	.31	.14
Student's Lang. (Coll. Lit.)		--	--	--	--	--	--	--	--	--	--	.08	-.04	.17	-.06	.01	.34	-.03
Hrs. Misc. Courses		--	--	--	--	.04	.11	-.31	.15	--	--	--	--	--	--	--	--	--
Canonical R		.61**	.67**	.53**	.59**	.65**	.68**	.49*	.63**	.67**	.77**	.67**	.70**	.66**	.62**	.73**	.60	.60**
p																	.10	
N		1039	289	80	764	1039	289	80	763	139	91	733	131	518	167	151	58	154
Criteria																		
Listening		.87	.91	.21	.67	.90	.96	.43	.73	.89	.42	.91	.78	.51	.80	.97	.59	.71
Speaking		.33	.40	.85	.51	.21	.27	.84	.29	.37	.80	.08	.58	.08	.57	-.07	-.01	.06
Reading		.03	.10	.47	.53	.21	.01	-.24	.59	.26	.35	.29	-.22	.47	-.15	.20	.25	.63
Writing		.37	-.09	-.10	.09	.33	-.05	-.23	.19	.12	-.27	.29	.02	.72	.13	.14	.76	-.31

**p < .01; *p < .05; significance levels given only for canonical correlation coefficients

Table 8.8

Canonical Regression Results for the WR (Superiority in Writing and Reading) Factor

Variable		Complete Samples, MLAT & Background				Complete Samples, MLAT, Background, & History				Grade School Starters		High School Starters			College Starters			
	Canonical:	II	II	II	II	II	II	IV	II	II	IV	II	II	II	II	II	I	III
	Language:	F	G	R	S	F	G	R	S	F	S	F	G	S	F	G	R	S
Predictors																		
MLAT-3		.10	-.19	-.13	.55	--	--	--	--	-.19	.13	.13	-.21	.14	.08	-.27	-.02	.34
MLAT-4		.85	.96	.15	.64	--	--	--	--	.48	-.34	.50	.87	.31	.61	.67	-.50	.28
MLAT-5		.23	-.12	-.52	.24	--	--	--	--	.22	.29	.28	-.19	.13	-.23	.05	.25	.18
MLAT-Total		--	--	--	--	.64	.62	-.59	.63	--	--	--	--	--	--	--	--	--
Time Began		-.07	-.02	-.66	.08	--	--	--	--	--	--	--	--	--	--	--	--	--
Time Abroad		-.46	.02	-.45	-.04	-.31	-.04	-.06	-.19	.14	-.10	-.43	.08	-.33	-.43	-.33	-.18	-.29
Use of FL at Home		-.05	.14	-.23	-.48	-.21	-.40	-.24	-.54	-.10	.00	-.15	-.13	-.51	.14	-.11	-.32	-.13
Sem. G. S.		--	--	--	--	.00	.02	.27	-.23	.08	-.13	--	--	--	--	--	--	--
Sem. H. S.		--	--	--	--	-.34	.18	.57	.22	.17	.38	-.06	.19	-.04	--	--	--	--
Teacher's Lang. (H. S.)		--	--	--	--	--	--	--	--	-.43	-.62	-.20	.03	-.11	--	--	--	--
Hrs. Coll. Lang.		--	--	--	--	-.37	.55	-.12	-.04	-.45	-.15	-.26	-.02	-.33	-.19	.27	.24	-.23
Teacher's Lang. (Coll. Lang.)		--	--	--	--	--	--	--	--	.17	-.43	.34	.28	-.21	-.21	.39	.29	.42
Student's Lang. (Coll. Lang.)		--	--	--	--	--	--	--	--	--	--	-.34	.03	.42	.07	-.31	-.44	-.49
Hrs. Coll. Lit.		--	--	--	--	.28	.23	-.37	.39	.44	.06	.23	.04	.03	.27	.13	.23	.31
Teacher's Lang. (Coll. Lit.)		--	--	--	--	--	--	--	--	-.11	.12	.00	.01	.33	-.14	-.01	.32	.28
Student's Lang. (Coll. Lit.)		--	--	--	--	--	--	--	--	--	--	-.23	-.11	-.20	.42	.13	.25	-.17
Hrs. Misc. Courses		--	--	--	--	-.34	.26	-.20	.11	--	--	--	--	--	--	--	--	--
Canonical R		.30**	.46**	.38	.27**	.31**	.35**	.25	.30**	.42*	.27	.38**	.51**	.29**	.36	.52**	.65*	.28
p				.46				.43			.68				.33			.46
N		1039	289	80	764	1039	289	80	763	139	91	733	131	518	167	151	58	154
Criteria																		
Listening		-.59	-.45	-.48	-.28	-.43	-.16	-.72	-.07	.04	-.77	-.62	-.37	-.46	-.76	-.33	.01	-.36
Speaking		-.35	-.44	-.46	-.44	-.54	-.66	-.02	-.61	-.84	.06	-.35	-.44	-.40	.05	-.44	-.66	-.63
Reading		.37	.21	.65	-.09	.64	.12	.59	-.15	.43	.50	.48	.11	-.04	.47	-.09	-.24	.35
Writing		.63	.75	.37	.84	.35	.72	.37	.77	.34	.40	.51	.81	.79	.44	.83	.71	.60

**p < .01; *p < .05; significance levels given only for canonical correlation coefficients

Table 8.9

Canonical Regression Results for the SW (Superiority in Productive Skills) Factor

Variable	Complete Samples, MLAT & Background				Complete Samples, MLAT, Background, & History				Grade School Starters		High School Starters			College Starters			
Canonical: Language:	III F	IV G	III R	IV S	III F	IV G	III R	IV S	III F	III S	III F	IV G	IV S	III F	IV G	IV R	II S
Predictors																	
MLAT-3	.05	-.62	-.63	.66	--	--	--	--	-.29	.13	.23	-.07	.24	.10	.37	-.12	.48
MLAT-4	.16	.30	.24	-.17	--	--	--	--	.57	.36	.06	.20	-.36	-.01	.18	.04	.05
MLAT-5	-.48	-.21	-.28	-.59	--	--	--	--	-.17	-.04	-.21	-.49	.15	-.27	-.44	-.29	-.29
MLAT-Total	--	--	--	--	.38	.19	-.31	.31	--	--	--	--	--	--	--	--	--
Time Began	-.50	.41	.53	-.28	--	--	--	--	--	--	--	--	--	--	--	--	--
Time Abroad	-.40	-.24	-.33	-.33	-.51	-.32	-.42	.00	.15	-.02	-.03	-.02	-.26	-.06	-.42	-.29	.09
Use of FL at Home	.58	.51	-.26	.04	.43	.59	-.18	.12	.12	.20	.60	-.01	.17	-.03	.46	-.04	-.08
Sem. G. S.	--	--	--	--	.25	-.03	-.06	.23	-.09	.28	--	--	--	--	--	--	--
Sem. H. S.	--	--	--	--	.28	.42	-.07	.49	.27	.42	.22	-.28	.50	--	--	--	--
Teacher's Lang. (H. S.)	--	--	--	--	--	--	--	--	.01	-.57	-.23	.03	.25	--	--	--	--
Hrs. Coll. Lang.	--	--	--	--	.14	.53	.71	.72	.37	.14	-.08	-.21	-.09	.24	-.19	.02	.68
Teacher's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	-.13	.22	.03	-.45	.17	.16	-.11	-.38	.05
Student's Lang. (Coll. Lang.)	--	--	--	--	--	--	--	--	--	--	.40	.38	.28	-.21	.07	-.08	.10
Hrs. Coll. Lit.	--	--	--	--	-.26	-.04	-.42	-.25	-.47	-.39	-.15	-.42	-.33	-.26	.17	-.63	-.32
Teacher's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	.27	.11	-.08	.26	.16	-.53	.41	-.09	.01
Student's Lang. (Coll. Lit.)	--	--	--	--	--	--	--	--	--	--	-.50	.00	-.36	.66	.04	.44	.30
Hrs. Misc. Courses	--	--	--	--	.43	-.23	-.06	-.12	--	--	--	--	--	--	--	--	--
Canonical R	.15**	.10	.19	.06	.18**	.09	.30	.13*	.39	.41	.21**	.30	.08	.24	.20	.31	.46**
P		.40	.91	.49		.78	.45		.15	.34		.32	.96	.91	.67	.72	
N	1039	289	80	764	1039	289	80	763	139	91	733	131	518	167	151	58	154
Criteria																	
Listening	-.60	.08	-.73	-.11	-.75	-.47	-.46	-.13	-.19	.07	-.69	-.45	-.11	-.50	-.55	-.83	-.23
Speaking	.59	.14	.27	.41	.49	.35	.30	.34	.03	.20	.71	.55	.44	.82	.61	.24	.14
Reading	-.34	-.80	-.13	-.75	-.06	-.46	-.60	-.72	-.56	-.78	.12	-.55	-.74	-.07	-.35	.19	-.52
Writing	.42	.58	.61	.51	.44	.66	.58	.59	.81	.59	.06	.44	.49	-.26	.45	.46	.18

Table 8.10

Canonical Regression Results for the RS (Superiority in Reading and Speaking) Factor

Variable	Canonical: / Language:	Complete Samples, MLAT & Background				Complete Samples, MLAT, Background, & History				Grade School Starters		High School Starters			College Starters			
		IV F	III G	IV R	III S	IV F	III G	II R	III S	IV F	II S	IV F	III G	III S	IV F	III G	III R	IV S
Predictors																		
MLAT-3		.68	.33	.36	.10	--	--	--	--	.45	.11	.16	-.05	.13	.20	.41	.05	-.32
MLAT-4		-.31	.22	.66	.27	--	--	--	--	.25	-.41	-.32	.16	.33	-.08	.46	.16	.30
MLAT-5		.05	-.70	-.40	.17	--	--	--	--	.05	-.37	.22	-.22	.10	-.18	-.36	-.08	.23
MLAT-Total		--	--	--	--	-.01	.55	.22	.60	--	--	--	--	--	--	--	--	--
Time Began		.53	-.45	.26	.42	--	--	--	--	--	--	--	--	--	--	--	--	--
Time Abroad		-.11	-.27	-.34	-.47	.02	.04	.04	-.21	-.17	-.35	.00	-.17	-.07	.02	.33	-.10	-.15
Use of FL at Home		.39	.29	.30	.70	.59	.05	.02	.26	.14	.29	-.11	.13	.47	.07	-.12	-.02	.29
Sem. G. S.		--	--	--	--	-.04	-.01	-.25	.06	.43	-.14	--	--	--	--	--	--	--
Sem. H. S.		--	--	--	--	-.75	.49	-.34	-.55	-.10	-.50	.21	-.02	-.33	--	--	--	--
Teacher's Lang. (H. S.)		--	--	--	--	--	--	--	--	-.16	.16	-.38	.11	.18	--	--	--	--
Hrs. Coll. Lang.		--	--	--	--	-.01	-.60	-.60	-.35	.36	.26	-.16	.14	-.47	-.24	-.49	-.32	.00
Teacher's Lang. (Coll. Lang.)		--	--	--	--	--	--	--	--	-.35	.13	-.48	-.34	-.22	.71	-.10	.06	-.35
Student's Lang. (Coll. Lang.)		--	--	--	--	--	--	--	--	--	--	.44	.23	-.14	-.42	-.09	.00	.13
Hrs. Coll. Lit.		--	--	--	--	.30	.20	-.64	-.17	.23	-.21	.09	.36	-.40	.26	.02	.12	-.24
Teacher's Lang. (Coll. Lit.)		--	--	--	--	--	--	--	--	-.39	-.24	-.30	-.64	.03	-.34	-.19	-.60	.67
Student's Lang. (Coll. Lit.)		--	--	--	--	--	--	--	--	--	--	.26	.36	-.21	-.04	-.27	.68	-.04
Hrs. Misc. Courses		--	--	--	--	.00	-.23	.08	-.26	--	--	--	--	--	--	--	--	--
Canonical R		.11**	.24**	.09	.14*	.06	.21	.43	.17**	.18	.49	.12	.36	.18	.08	.35	.48	.19
	P			.90		.68	.24	.17		.89	.10	.43	.17	.53	.99	.13	.44	.68
	N	1039	289	80	764	1039	289	80	763	139	91	733	131	518	167	151	58	154
Criteria																		
Listening		-.30	-.77	-.50	-.79	-.15	-.62	-.43	-.82	-.75	-.59	.02	-.60	-.73	.09	-.54	-.39	-.71
Speaking		.30	.58	.33	.35	.28	.18	.47	.23	.24	.58	.06	.25	.40	.02	.06	-.33	.51
Reading		.69	.28	.60	.50	.64	.74	.41	.49	.62	.28	.68	.70	.54	.68	.80	.85	.47
Writing		-.59	-.01	-.53	.01	-.70	-.20	-.65	.18	.02	-.48	-.73	-.28	-.11	-.73	-.24	-.13	-.14

them were statistically significant at the 5% level or better. Many, indeed, were at a purely chance level. There are, in fact, only 3 ways of dividing 4 things into 2 sets of 2 each, and this factor may represent little more than a chance residual from other sources of variance. It is a little hard to imagine just what superiority in reading and speaking (as contrasted to listening and writing) would imply.

Let us now consider the results shown in Table 8.7. The canonical regression weights for particular combinations of predictors are shown in the first part of the table, followed by the values of the canonical correlations, probabilities (given only where they are greater than .05), and numbers of cases (N's). The data of this table generally match those of Tables 8.2 through 8.5 because this table, like those, has to do with measures of overall competence in the relevant foreign language. The almost universally best predictor of measured competence is the Amount of Time Spent Abroad, and this finding agrees with those noted elsewhere in this study. The results shown here emphasize the fact that Time Abroad is particularly associated with Listening, and also (though to a lesser extent) with Speaking.

When three separate MLAT subtests are included in the analysis, tests 3 and 4 are more likely to show positive coefficients for predicting overall language competence. In the second set of data (under the heading "Complete samples, MLAT, Background, and History") the MLAT Total shows all positive and sometimes substantial coefficients.

Time of Beginning FL Study shows negative coefficients that reflect the fact that those who started earlier are likely to achieve higher skill than those who start late.

The coefficients for Use of the FL at Home are in the expected direction and are usually of substantial magnitude, showing that this variable is an important influence, independent of the other variables, in the attainment of overall skill.

The data for various indices of length and quality of exposure to formal schooling in the foreign language are generally such as to suggest that this formal training is influential in enhancing foreign language skill. The coefficients for the indices relating to college literature courses are often sizable, although they are never as large as those for Time Abroad. Only in the college-starting group in Spanish do we find a sizable coefficient for the influence of language competence of the teacher in a college language course. Some of the coefficients for language courses are negative, especially for groups who should have been able to complete introductory language courses at an earlier stage.

Table 8.8, it will be recalled, has to do with a dimension of language competence whereby the individual shows significantly different skills in reading and writing as opposed to his skills in speaking and listening. We have chosen the direction of this "bipolar" scale so that positive coefficients imply a superiority of reading and writing over the listening and speaking skills. The reason for choosing this direction will now be evident: the best predictor of this dimension is generally subtest 4 of the MLAT (or, where that subtest has not been separately entered into the analysis, the total score from the 3 MLAT subtests). Subtest 4 of the MLAT, called Words in Sentences, is chiefly a measure of the individual's "grammatical sensitivity," i.e., his ability to identify and relate the grammatical functions of words in sentences. It is not surprising that such an ability might be associated with the individual's superiority in the written aspects of his foreign language attainments, particularly as measured by the Writing subtest of the MLA battery, since the Writing test probably puts the student on his grammatical mettle more than any other of the MLA skills tests.

There does exist, however, an apparent anomaly in these results, in that the MLAT was originally designed and validated to predict rate of learning in audiolingual skills. Why does it now show better prediction of reading and writing skills? In the first place, it must be noted that only the Words in Sentences subtest (MLAT-4) shows the generally anomalous results; frequently the other subtests have negative coefficients for the writing superiority factor WR (and thus, in effect, are positively associated with audiolingual superiority). Secondly, the students contained in our samples are presumably ones who have had, in many cases, many years of exposure to foreign language training. The MLAT was designed to measure mainly rate of learning in introductory language training. The students in our samples have generally long since passed the stage of introductory language training and have had an opportunity to overcome any effects of deficient general language aptitude; that is, those of lower aptitude, while taking perhaps longer than those of higher aptitude, have nevertheless had sufficient time to achieve minimal competence in their major language.

A variable that has generally negative coefficients in predicting superiority of reading and writing over audiolingual skills is Amount of Time Abroad. These values could indicate simply that Time Abroad is positively associated with superiority in listening and speaking skills; it is also possible, however, that those who "stay at home" would have more opportunity to cultivate their reading and writing skills.

The findings for Use of the FL at Home indicate, similarly, that exposure to the foreign language at home is associated with superiority of audiolingual skills.

The results for the several variables reflecting length and quality of formal foreign language training are none too consistent over the various samples and language groups. Variables having to do with college literature courses do, however, generally have positive coefficients suggesting that such courses promote reading and writing skills more than they do audiolingual skills. Conversely, variables derived from experiences in introductory and intermediate language courses tend to have negative coefficients, suggesting that they promote audiolingual skills. The patterns of coefficients across languages suggest that French and Spanish language courses tend to concentrate on audiolingual skills while German and Russian language courses emphasize the promotion of reading and writing skills. Nevertheless, even in German and Russian, the results suggest that students with experience in classes where they are required to use the foreign language tend to gain in audiolingual skills more than in reading and writing skills.

A possible dimension of superiority in productive skills (speaking and writing), the SW factor, is dealt with by Table 8.9. None of the canonical correlations are as high as those generally found in Table 8.7 or 8.8, and the patterns of canonical regression weights for the predictors are not very consistent across samples or languages. The language aptitude subtests show a very puzzling pattern of results: MLAT-4 has generally positive coefficients, MLAT-5 generally negative coefficients, and MLAT-3 sometimes positive and sometimes negative coefficients. It does not seem worthwhile to speculate about these differences.

The coefficients for Time of Beginning suggest that in French and Spanish, those who begin early are more likely to attain superiority in productive skills, while in German and Russian early beginning is more often associated with superiority in receptive skills. These differences may reflect differences in the ways in which these languages are typically taught, German and Russian more often being taught with a reading comprehension objective than with a production mastery objective.

Somewhat surprisingly, Amount of Time Abroad is more associated with superiority in receptive than in productive skills. At the levels of mastery typically attained by students in our sample, Time Abroad may have more effect in enriching the receptive experience of the students than in promoting active language skills.

In contrast, Use of the FL at Home is more likely to be positively associated with the development of productive skills, according to the data. Students with exposure to the foreign language at home are perhaps more likely to be at the level of mastery requisite for them to profit from this experience in the direction of developing productive skills.

The coefficients for variables derived from data on the formal training of the students suggest that introductory and intermediate courses tend to promote productive skills, while literature courses promote receptive skills. These findings square with expectations. It is noteworthy, however, that for high school starters, active use of the language by students in college literature courses promotes receptive skills, while for college starters, it promotes productive skills.

It is difficult to interpret the results for the RS factor, shown in Table 8.10. Even if we restrict attention to the few canonical correlations that are significant at the 1% level, the patterns of coefficients for the predictors are highly inconsistent. Perhaps Table 8.10 is best left as an exhibit on which the reader can test his imagination.

6. The stability of regression systems over institution size strata and institution types

In Chapter V, it was shown that there were large differences among strata of institution-sizes and between institution types (public and private) in variables having to do with student backgrounds and with instructional procedures. There was concern over whether institutional differences in mean MLA skills test scores might be largely due to such differences. Even though it was shown that the mean MLA skills test score differences remained even after certain statistical controls were applied, there was still some doubt as to whether problems of institutional differences had been adequately resolved.

A further attempt was made to resolve these problems by applying the canonical correlation technique *within* strata and institution types. It was reasoned that if the first canonical correlations were found to be drastically lower in such analyses than they were in tabulations done across strata and institution types, the conclusion could be drawn that identifiable stratum and institution-type differences were largely responsible for the variation in mean MLA skills test scores. This conclusion would tend to cast doubt on the success of the sampling design in drawing a representative sample of institutions. If, on the other hand, the canonical correlations remained high, one would conclude that sufficient variation existed *within* cells of the sampling design to permit one to discount between-cell differences as being important sources of variance.

Table 8.11 presents the results of this analysis. It was performed only for French because this was the only language with enough cases to permit sub-classifications with substantial numbers of cases. (Even so, students in Public Stratum 1 were not numerous enough to justify application of the canonical correlation technique.) The table presents data only for the first canonical correlation. As may be seen in the table, the canonical correlations were fully as high as, if not higher than, those obtained for the total sample in French. In general, the patterns of large coefficients are similar across strata and institution types. Time Abroad is nearly always the best predictor; the exception is in Public Stratum 4, and this is possibly because the criterion composite seems to represent not overall competence but the SW factor (the contrast between Speaking and Writing, on the one hand, and receptive skills of Listening and Reading). In any case, the correlation is barely significant. Time Began is generally a negative predictor, as it has been found to be for the French sample as a whole. Hours of College Literature, or the instructor's use of the FL in college literature courses, is a reasonably consistent predictor.

It may be concluded, we feel, that within-cell variations in student background and experiences are of sufficient importance to allow us to discount the importance of between-cell differences in accounting for the results of this study.

Table 8.11

First Canonical Regression Weights, by Stratum and Institution Type, for 14 Selected Predictor Variables and MLA Skills Tests, French "Regular" Cases

	Stratum 1 (Small)		Stratum 2		Stratum 3		Stratum 4		Stratum 5 (Large)	
	Pub.	Priv.	Pub.	Priv.	Pub.	Priv.	Pub.	Priv.	Pub.	Priv.
Predictor Variables:										
Time Began	not significant	-.03	.20	-.34	-.21	-.17	-.16	-.29	-.16	-.22
Time Abroad		.56	.38	.57	.53	.87	.08	.69	.64	.45
Use of FL at Home		.38	.17	.11	.26	.01	.01	.07	.36	-.02
Hrs. Coll. Lang.		.11	-.14	.36	-.01	-.15	.33	-.29	-.14	-.32
Teacher's Lang. (Coll. Lang.)		-.05	.47	.30	-.10	.27	-.09	-.05	.06	.38
Student's Lang. (Coll. Lang.)		.29	-.33	-.16	.06	-.16	-.06	.33	.26	-.14
Use of Lang. Lab. (Coll. Lang.)		-.02	.14	-.22	-.15	.02	-.04	.07	-.04	.42
Teacher's accent (Coll. Lang.)		-.30	.16	-.10	.26	.02	.17	-.26	-.09	-.25
Hrs. Coll. Lit.		.43	.49	.13	-.06	.19	.44	.01	.38	.37
Teacher's Lang. (Coll. Lit.)		.00	.08	.38	.57	.23	.53	.34	-.35	.15
Student's Lang. (Coll. Lit.)		.25	.12	-.22	-.31	.05	-.37	.19	.20	.10
Use of Lang. Lab. (Coll. Lit.)		.16	-.33	-.18	-.27	-.06	-.20	-.01	-.09	-.17
Teacher's accent (Coll. Lit.)		.27	-.06	-.03	-.08	-.07	-.37	.13	.06	.08
Hrs. Misc. Courses		.01	-.13	.05	.08	.05	-.16	.04	.17	.19
Canonical R	--	.81**	.77**	.69**	.74**	.70**	.69	.69**	.65**	.60**
P							.06			
N	27	84	51	111	61	198	66	188	203	111
Criteria:										
MLA Listening	--	.85	.86	.67	.43	.87	.81	.94	.82	.88
MLA Speaking	--	.52	.41	.61	.46	.14	-.23	.18	.55	.27
MLA Reading	--	.03	.15	.26	-.46	-.13	.39	-.22	-.09	.39
MLA Writing	--	.00	.26	-.34	.63	.46	-.36	.17	.08	.04

Chapter IX

CHARACTERISTICS OF MODERN FOREIGN LANGUAGE DEPARTMENTS IN RELATION TO STUDENT ACHIEVEMENT

1. Introduction

Chapter V showed that certain characteristics of institutions--namely, institution size and type (public vs. private)--were related to student achievement. Its findings left the impression that student achievement might be related to variations in instructional policies and procedures.

Besides data collected from the students themselves concerning the characteristics of the instruction they had had, information was obtained by means of the Questionnaire for Chairmen of Foreign Language Departments (Appendix C), completed and returned by at least one of the foreign language departments in 198 of the institutions participating in the national study. In all, 334 questionnaires were returned. The number of questionnaires returned from a given institution depended partly on the number of separate foreign language departments at the institution and partly on the willingness of the department chairmen to complete and return the questionnaire. According to our records, the 334 questionnaires returned represent 91.0% of the total number, 367, of separate departments in the 202 participating institutions. This is a high rate of return.

The analysis of the data contained in these questionnaires has been reported in a doctoral thesis by Fannie A. Handrick (Handrick, 1967). Data from both the pilot study sample in the state of New York and the national sample were studied. Since this thesis is (or will be) available both through interlibrary loan and also through microfilming services (including the ERIC system of the U. S. Office of Education), we present here only a summary of the major findings; we restrict this summary to data from the national sample.

The unit of analysis was sometimes the "department" and sometimes the "subdepartment," defined as a department or part thereof devoted to the teaching of one of the five languages included in this study. In this sense, the 334 departments represent 769 subdepartments. Table 9.1 shows a classification of these subdepartments by department type, institution size, and institution type. "Department type" is indicated either as "single," "family," or "all." "Single" means that the department was devoted to only one language (e.g. Department of French); "family" means that the department was devoted to a family of languages (e.g. Department of Romance Languages); "all" means that the department embraced all the foreign language teaching (at least, modern languages) that was done at the institution. As one might expect, subdepartments responsible for instruction in a single language are most often found in large, private institutions.

The assumption was made that the instructional policies and procedures reported by the department chairman (or whoever filled out the questionnaire) applied to any or all of the language instruction covered by the department. For example, when the subdepartment was the unit of analysis, the data from a questionnaire returned by a department of Romance Languages was used for the analysis of both the French and the Spanish instruction.

2. Dimensions of Instruction in Subdepartments

Factor analytic techniques were applied to the questionnaire data to identify the chief dimensions in which instructional policies and procedures in beginning and intermediate courses varied. There were four such dimensions identified:

I. The extent to which the use of English is avoided in first year courses (i.e., introductory language courses). This dimension reflected chairmen's responses to a ques-

Table 9.1

National Study: Classification of Sub-Departments by Type of Department and Type and Size of Institution

	French N = 189	German N = 178	Italian N = 64	Russian N = 129	Spanish N = 189	Total N = 769
Department Type						
Single	33	37	8	12	32	122
Family	40	28	37	17	40	162
All	116	113	39	100	117	485
Institution Size						
1 (small)	25	20	4	7	25	81
2	28	25	5	17	28	103
3	60	5	18	36	58	227
4	40	39	25	34	41	179
5 (large)	36	39	32	35	37	179
Institution Type						
Private	121	110	44	69	121	465
Public	68	68	40	60	68	304

tion that asked how often English was used to explain points of grammar or the meanings of new words and phrases. The major contrast was between those departments that reported English was used "frequently" for these purposes and those departments reporting "occasional" use of English.

II. The extent to which the use of English is avoided in second-year (intermediate) courses. Here the major contrast was between departments reporting "occasional" use of English and "rare" use of English. There was some tendency for these two dimensions to correlate; that is, departments reporting "occasional" use of English in the first year were more likely to be those departments reporting "rare" use of English in the second year; likewise, departments reporting "frequent" use of English in the first year were more likely to be departments reporting "occasional" use in the second year.

III. The emphasis given to audiolingual objectives. The major contrast was between those departments reporting "more emphasis on listening and speaking" and those departments reporting "equal emphasis given to all four skills." (Very few departments reported that they gave more emphasis to reading and writing than to listening and speaking.)

IV. The opportunity given for voluntary use of a language laboratory. The major contrast was between departments requiring use of the language laboratory in basic skills courses and those departments allowing purely voluntary use of the language laboratory.

Factor scores based on these dimensions were assigned to 656 subdepartments with requisite data, with a view to relating these departmental characteristics to data on student achievement.

There was some question, however, as to whether the data based on the reports of department chairmen corresponded to student reports concerning what went on in the classroom. Information was assembled from the reports of students who started study of their major foreign language at the college level. Ratings of first and second year courses listed by these students in the Questionnaire for Foreign Language Majors (Appendix B) were averaged within subdepartments. By using data only from subdepartments where there were at least two students contributing ratings, it was possible to estimate the average reliability of these ratings. The reliabilities so estimated are as follows:

	Reliability
Teacher's use of English, first year courses	.39
Teacher's use of English, second year courses	.63
Audiolingual emphasis	.66
Use of language laboratory	.67

When these ratings were compared with the factor scores assigned to the departments, the correlations were significant but very low, as shown by Table 9.2.

Table 9.2

National Study: Correlations Between Four Factors of Audiolinguality and Mean Student Ratings of Classroom Practices

Factor	Student Rating	N	r
First year classroom use of English	Teacher's language, first year	245	.19**
Second year classroom use of English	Teacher's language, second year	244	.20**
Audiolingual emphasis	Audiolingual emphasis	243	.13*
Opportunity for voluntary use of language laboratory	Use of language laboratory	243	-.06

**p < .01
*p < .05

It should be noted, however, that the questions asked of the students and of the department chairmen were not precisely comparable. Highest agreement is found when the questions were essentially the same, viz. to what extent does the teacher speak English in the classroom? The questions concerning audiolingual emphasis and use of the language laboratory asked of students and of department chairmen were less similar. Nevertheless, the overall amount of agreement between students' and department chairmen's ratings was not great enough to justify thinking that the two sets of ratings referred to the same things, and this fact must be borne in mind as the data are examined.

Subdepartments were also scored for "advanced instruction variables": number of degree requirements, number of required and voluntary activities offered (opportunities for study and travel abroad, language houses, clubs, a pledge to speak only the foreign language at designated places and times), and numbers and types of courses offered. A summary of the data bearing on advanced instruction is given in Table 9.3. These data may be summarized conveniently by the statement that subdepartments teaching the more popular languages (French, Spanish, German) have more degree requirements, offer more associated language-learning activities, and offer more courses than subdepartments concerned with the less popular languages (Italian, Russian). These advanced instruction variables are also related to institution size: larger institutions tend to score higher on them than smaller ones do. Furthermore, public and private institutions differ in these advanced instruction variables: public institutions, while offering more types of courses, tend to report fewer degree requirements and associated language-learning activities than private institutions do.

Table 9.3

National Study: Advanced Instruction Variables

A. Percentages of Sub-departments Mentioning Each Type of Degree Requirement

Requirements:	French N = 189	Spanish N = 189	Italian N = 84	German N = 178	Russian N = 129
Advanced Skills	85.8	81.1	27.4	70.6	45.4
Literature	78.0	72.8	27.4	62.7	41.6
Culture & Civ.	36.9	32.8	7.1	28.0	16.9
Linguistics	8.8	9.9	3.6	9.5	6.2
Written test	16.6	16.6	2.4	10.6	3.9
Oral test	3.6	4.2	1.2	3.4	.8

B. Percentages of Sub-departments Requiring Various Types of Activities

Activities:	French N = 189	Spanish N = 189	Italian N = 84	German N = 178	Russian N = 129
Year abroad	--	--	--	0.6	--
Summer abroad	0.5	0.5	--	0.6	--
Language house (school year)	1.6	1.0	--	0.6	--
Language house (summer)	--	--	--	--	--
Language table	2.1	2.6	--	1.7	2.3
Language club	5.0	3.6	--	3.4	0.8
Pledge	1.0	1.0	--	0.6	1.5

C. Percentages of Sub-departments Offering Various Activities on a Voluntary Basis

Activities:	French N = 189	Spanish N = 189	Italian N = 84	German N = 178	Russian N = 129
Year abroad	56.7	49.4	26.2	48.2	10.0
Summer abroad	53.6	53.6	21.4	43.1	19.3
Language house (school year)	15.1	8.3	--	7.3	4.6
Language house (summer)	5.7	3.6	1.0	3.4	3.9
Language table	41.6	33.8	10.7	40.9	20.8
Language club	89.2	70.2	42.8	63.8	46.2
Pledge	5.0	5.0	--	4.5	6.9

D. Percentages of Sub-departments Offering Various Types of Courses:

Types of Courses:	French N = 189	Spanish N = 189	Italian N = 84	German N = 178	Russian N = 129
Beginning	93.6	95.2	95.2	96.8	96.3
Intermediate	96.2	94.6	84.5	96.3	92.4
Composition	88.9	86.3	38.1	80.1	53.9
Conversation	86.3	85.3	32.1	80.1	54.7
Phonetics	--	--	--	--	--
Advanced Grammar	54.6	52.0	13.1	36.4	29.3
Literature	95.2	90.5	55.9	90.2	64.7
Special Reading	36.9	26.0	13.1	57.1	42.4
Culture & Civ.	61.9	56.2	15.5	47.0	29.3
Diachronic Linguistics	21.3	18.7	14.3	25.8	20.0
Synchronic Linguistics	25.0	25.0	11.9	23.0	23.1

3. Departmental characteristics and student achievement

Complete data on departmental characteristics, mean student aptitude, and mean student achievement were available for 163 of the 202 institutions participating in the national study. These institutions represented 383 sub-departments of foreign languages apportioned among the five languages as follows: French, 133; German, 86; Italian, 10; Russian, 28; and Spanish, 126. Because of the small numbers of subdepartments in Italian and Russian, data are not analysed for these languages.

In computing mean student aptitude (from the MLAT total score) and mean student achievement (from the four MLA skills tests), only data from "regular" cases were used (see Chapter III). Two sets of subdepartment scores were computed. The first was derived from the mean scores of students who started their major foreign language study in college; the second was derived from the mean scores of all students majoring in the subdepartment, regardless of when they started their foreign language study. The former scores were derived because it was reasoned that data on subdepartment policies for beginning and intermediate language courses would be relevant only for students who would have taken such courses, namely, those starting in language at the college level. The latter set of scores, however, would be relevant for all students majoring in the subdepartment, because they could conceivably have been affected by policies and procedures in advanced language instruction. The means were computed over the students actually in a subdepartment, even if the basis for such a mean was only one student, as was sometimes the case. The numbers of students involved in each set of scores, over all subdepartments, are shown in Table 9.4.

Table 9.4

Numbers of Students Tested in Each of Two Groups by Language

Language	Started in College	All Students
French	191	1140
German	174	340
Italian	14	25
Russian	73	97
Spanish	167	833
	619	2435

Unfortunately, the reliability of mean scores computed in this way tended to be low, particularly for scores based on students starting in college. To estimate the reliability coefficients for such mean scores, the intraclass correlation was obtained (Ebel, 1951) for means based on 2 or more students within a subdepartment: when these reliabilities were found to be low, the computations were restricted to subdepartments having at least 3, or in some cases 4, students, but such restriction reduced the number of subdepartments drastically.

For students starting their major language in college, only in Spanish was it found that mean aptitude and MLA skills test scores were sufficiently reliable to constitute variables that could be used in studies of the relation between departmental variables and mean student aptitude and language achievement. The intraclass correlations based on subdepartments with 2 or more college-starting students ranged from .92 for Writing to .96 for Listening and for Speaking, in the case of the MLA skills tests; the intraclass correlation for mean student aptitude was .78. (High reliability signifies, essentially, that variation between subdepartments was significantly greater than it was within subdepartments.

A regression analysis based on 70 institutions with college-starting students tested in Spanish is summarized in Table 9.5. The multiple correlations are "shrunken" to correct for sample size. The purpose was to see whether departmental characteristics, as reported either by the department chairmen or by the students themselves, could predict mean student achievement on the MLA Skills tests. Actually, institution size was the only predictor whose beta-weights were more or less consistently significant; this finding only repeats that already made in Chapter V. Because of the small sample of institutions, we cannot rely on the other beta-weights. There is a trend, however, for the variable called Avoidance of English in First Year Courses (as reported by department chairmen) to show positive beta-weights. (The corresponding zero-order correlations were .28, .04, .23, and .24 for the four MLA skills tests respectively, a correlation as large as .23 being significant at the 5% level.) None of the other factor scores for departmental characteristics related to beginning and intermediate language instruction showed significant trends, except that if anything, departments reporting emphasizing listening and speaking skills tended to show slightly lower mean achievement in their students than departments reporting all four skills emphasized equally. Among the variables derived from student reports, those related to the teacher's avoidance of English tended to have positive beta weights, in at least one case significant at the 5% level. Specifically, subdepartments in which students tended to report that the instructors in their second-year language classes avoided the use of English tended to be departments that had somewhat superior student mean scores on the MLA skills test; the corresponding simple correlations were .27, .18, .45, and .36 for Listening, Speaking, Reading and Writing, respectively.

The validity coefficients of the student ratings are, in general, higher than those of the factor scores derived from department chairmen's questionnaire responses. Apparently, students are in a better position to know what goes on in classrooms than department chairmen are. On the other hand, the effects reported here are very small in magnitude; undoubtedly they are overridden by more important variables such as the amount of time abroad and others that we have examined in previous chapters.

Handrick has also reported regression systems based on data from all students within a department regardless of their time of starting study of the major foreign language. The scores representing mean student achievement in MLA skills tests within subdepartments are in general quite reliable, showing that departments differ significantly, for whatever reasons, in the average performance of their students on the MLA skills tests. Data were computed for 133 French subdepartments, 86 German subdepartments, and 126 Spanish subdepartments. We do not consider it worthwhile to reproduce the results in detail, because they are largely negative and/or uninformative. Either institution size or department type is generally the only significant predictor of mean student MLA skills test scores. That is to say, single-language departments, and departments in larger institutions, tend to have students superior to those of "family" or multiple-language departments, or departments in smaller institutions. In some of the analyses, however, certain instructional policies have a significant but slight relationship to mean student performance. For example, in French, the extent to which English is to be avoided in first year courses is correlated significantly with mean student performance on Listening, and to some extent with Speaking. (It is to be remembered, however, that the student performance means were computed for all students in the department, the majority of whom must have started French in high school.) In German, this same variable is significantly related to student performance in Listening and Speaking, but not in Reading or Writing, and in Spanish it is significantly related to mean performance on Listening, Speaking, and Writing. One wonders, however, how a departmental policy specified in terms of avoidance of English in first-year courses could be related to mean performance of students most of whom do not take those courses. It is also curious that avoidance of English in second-year courses is in no case significantly related to mean performance of all students in the department. It can only be speculated that a policy or practice of avoiding English in first-year courses reflects a general departmental policy that also manifests itself in the more advanced courses. Unfortunately, the department chairmen were not asked about policies and practices in teaching the more advanced courses. If they had been, this study might have obtained further confirmation of findings in earlier chapters that sug-

Table 9.5

Multiple Regression Data for Prediction of Mean Student Achievement from Selected Departmental Variables
N = 70 Spanish Subdepartments,
Variables derived from Students who Started
Study of Spanish in College
(from Handrick, 1967)

	MLA Skills Test							
	Listening		Speaking		Reading		Writing	
Variable	r	β	r	β	r	β	r	β
Mean MLAT	-.09	-.07	-.15	-.07	.07	.07	-.06	-.01
Department Type[1]	-.31*	-.17	-.03	.08	-.33*	-.21	-.33*	-.22
Institution Size[1]	.36**	.36*	.20	.30*	.35*	.29*	.39*	.19
Institution Type[1]	-.11	-.29*	-.03	-.12	.07	.06	.14	.06
Department Chairmen's Data:								
Avoid English (1st year courses)	.28*	.18	.04	.03	.23*	.09	.24*	.19
Avoid English (2nd year courses)	.04	.02	-.04	-.07	.16	.19	.03	.04
Audiolingual Emphasis	-.05	-.14	-.19	-.15	-.01	-.04	-.03	-.11
Opportunities for volunteer lab. use	-.03	-.06	-.06	-.15	.15	.13	-.03	-.12
Degree Requirements	.01	-.07	-.04	.04	-.02	-.24*	-.07	-.11
No. of Activities	-.04	-.09	-.08	-.12	-.01	.09	-.14	-.14
Types of courses offered	.10	.06	-.08	-.12	-.06	-.18	.12	.02
Mean Student Ratings:								
Avoid English (1st year courses)	.14	.15	-.13	-.19	.24*	.20	.12	-.04
Avoid English (2nd year courses)	.27*	.15	.18	.18	.45*	.28*	.36*	.22
Audiolingual Emphasis	.12	-.10	.04	.01	.27*	-.12	.25*	-.15
Language lab. use	-.07	-.12	.09	.03	.10	-.02	-.05	-.18
Multiple R (shrunken)	.35		(.00)		.43		.37	

**p < .01; *p < .05 (two-tailed test)

[1] Coding:
Department Type: 1, Single language; 2, Family of languages; 3, All languages
Institution Size: 1, Small; . . ., 5, Large
Institution Type: 1, Public; 2, Private

gested that emphasis on speaking the foreign language in advanced language and literature courses was beneficial in enhancing language skills of the students.

There was no support in these analyses for the notion that use of a language laboratory makes any critical difference in student performance. We have already noted that it is only the smaller institutions, that may not have sufficiently competent teachers, that seem to emphasize the use of the language laboratory at the college level. This is not to say that the use of a language laboratory in larger institutions is pointless, but as far as our results are concerned, it does not seem to make any special contribution to improving student performance over and above what may be contributed by teachers who are competent in the language they are teaching and who emphasize its use in the classroom.

On the whole, these analyses of the relation of departmental characteristics to student performance leave many questions unanswered. It is probable that the Questionnaire for Foreign Language Department Chairmen failed to ask the right questions about instructional practices and policies--questions that would illuminate the differences between "good" departments and "poor" departments. The fact that our analyses showed reliable differences among departments in mean performance of their students presents a challenge to future researchers.

Chapter X

RESULTS FROM THE MLA TEACHER PREPARATION TESTS

1. The Sample

Coordinators at participating institutions were told (Appendix D-4): "The COMPLETE Battery of 7 tests [constituting the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students] may be administered only to those senior class students who are enrolled at your institution in a program which is expressly oriented toward the preparation of elementary or secondary school language teachers, and which usually involves the completion of a certain series of courses necessary for state certification purposes."

Of the 2874 students who took any of the MLA tests, 1116 (38.8%) took one or more of the Teacher Preparation subtests, tests 5, 6, and 7, and may therefore be presumed to have been enrolled in teacher certification programs. No information has been developed as to whether those enrolled in such programs were more likely than others to volunteer to participate in the testing program, or whether, therefore, they biased the sample to any extent. Students enrolled in teacher preparation programs may have had a special incentive to participate since they might have expected to have to take the MLA battery at some future time as a part of teacher certification procedures or as an examination to qualify for a teaching position; the opportunity to take the MLA tests free of charge and for practice only should have had considerable appeal. In institutions in the State of Pennsylvania, the results of the testing done in this project were, however, used as a part of regular certification procedures, by arrangement with state authorities; it is therefore likely that a greater proportion of teacher preparation candidates took the tests in that state than in other states, since they were apprised of this arrangement.

After the elimination of "odd" cases (see Chapter III), there remained approximately 1000 "regular" cases who had taken one or more of the Teacher Preparation tests of the MLA battery. Norms based on these cases have already been given in Table 4.3. A recapitulation of the norm data is given in Table 10.1, together with comparative data drawn from various publications of Educational Testing Service concerning the performance of NDEA Institute Teachers on these tests. Typically, the means of the CFLT samples are midway between NDEA pretest and posttest means on each of the three tests. The foreign language majors were better prepared, on the whole, than teachers arriving at NDEA Institutes. It is perhaps too much to expect the CFLT samples to be able to surpass the performance of the NDEA teachers after the latter had had the special training afforded them in the NDEA Language Institutes.

For most of the analyses presented in this chapter, the samples used included both "regular" and "odd" cases in French, German, Russian, and Spanish whose data were complete on all seven tests, i.e., both the teacher preparation tests and the skills tests. There were 985 such students.

As contrasted with students who took none of the Teacher Preparation tests, students who did take the tests tended to be somewhat inferior in MLA skills test performance. The comparisons are shown in Table 10.2, where it may be seen that mean MLA skills test scores of French and Spanish teacher preparation cases were significantly below those of French and Spanish students who did not take these additional tests. The comparisons of means were not significant, however, for the German and Russian groups. The results for French and Spanish groups agree with findings in Chapter VI where it was shown that students intending to teach in elementary or secondary school tended to have somewhat lower MLA skills test scores than those intending to teach at the college level, or not intending to teach.

There was also a tendency for the CFLT samples to be more homogeneous than the NDEA samples on the Teacher Preparation tests, that is, the standard deviations of the scores were in nearly every case smaller. (The exceptions were the comparisons with the standard

Table 10.1

Summary Statistics for MLA Teacher Preparation Tests, Norm Groups (From Table 4.3), compared with Data from NDEA Institute Standardization Groups[1] (ETS, 1961a, 1961b, 1962a, 1962b, 1962c, 1962d, 1964a)

		MLA Teacher Preparation Tests								
		Applied Linguistics (MLA-5)			Cult. & Civ. (MLA-6)			Prof. Preparation (MLA-7)		
Language	Group	N	$\bar{X}$	σ	N	$\bar{X}$	σ	N	$\bar{X}$	σ
FRENCH	CFLT	445	48.66	7.23	460	46.10	6.46	435[2]	62.77	7.11
	NDEA Pretest	c.700	45.46	8.40	c.700	44.33	8.95	1336	59.1	7.9
	NDEA Posttest	c.700	50.70	8.42	c.700	50.15	8.94	1336	66.9	6.7
GERMAN	CFLT	116	49.88	7.28	111	50.93	8.75	128[2]	62.52	7.19
	NDEA Pretest	c.150	48.04	9.68	c.150	48.62	10.11	297	60.2	7.6
	NDEA Posttest	c.150	54.64	9.30	c.150	53.97	9.06	297	67.3	6.2
ITALIAN	CFLT	8	49.38	4.95	8	46.12	6.45	(Not computed)		
	NDEA Pretest	c.65	46.1	6.5	c.65	49.8	8.0	40	53.9	6.2
	NDEA Posttest	c.65	50.7	7.8	c.65	54.5	8.5	40	61.9	6.7
RUSSIAN	CFLT	18	49.11	6.25	20	52.80	8.17	23[2]	60.04	5.26
	NDEA Pretest	c.110	44.25	6.79	c.110	49.39	9.15	176	59.7	8.5
	NDEA Posttest	c.110	48.45	7.84	c.110	53.42	8.86	176	64.4	7.0
SPANISH	CFLT	394	46.68	7.43	405	55.54	7.66	399[2]	61.17	7.83
	NDEA Pretest	c.700	43.07	7.69	c.700	49.53	9.82	1334	58.8	8.1
	NDEA Posttest	c.700	48.42	8.51	c.700	55.95	9.42	1334	66.8	6.4
ALL	CFLT							991	62.13	7.43
	NDEA Pretest							1745	59.32	8.10
	NDEA Posttest							1709	66.25	6.81
Totals	CFLT	981			1004			991		

[1] NDEA data on Form A (then labelled JML-1) are given, except for MLA-7 in the specific language groups, where the data are from both Form A and Form B combined because of the spiraling of these forms through institutes and through pretest and posttest administrations.

[2] These cases were those used in certain other tabulations and may include some cases that for some reason were not used in the norm groups. They were complete on all seven MLA tests.

Table 10.2

Comparisons on MLA Skills Tests of Those Who Did and Did Not Take Any Teacher Preparation Tests

			MLA Skills Test Scores							
			Listening		Speaking		Reading		Writing	
	N	%	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ
FRENCH MAJORS										
Took MLA 5-7	435	32.2	41.98	7.68	72.80	9.23	46.32	8.43	45.54	9.63
Took none of these	735	62.8	45.68	7.24	74.59	9.64	50.58	8. 6	48.50	9.50
Total	1170	100.0	44.31	7.63	73.92	9.54	49.00	8.64	47.39	9.66
t			-8.2**		-3.1**		-8,4**		-5.1**	
GERMAN MAJORS										
Took MLA 5-7	128	33.0	44.26	8.90	90.03	15.62	52.08	10.47	54.42	13.51
Took none of these	260	67.0	45.07	8.55	89.62	13.77	52.97	9.88	54.26	13.57
Total	388	100.0	44.80	8.69	89.76	14.42	52.68	10.10	54.31	13.57
t			-0.9		0.3		-0.8		0.1	
RUSSIAN MAJORS										
Took MLA 5-7	23	22.5	43.57	7.16	85.74	16.17	44.87	12.33	62.57	11.89
Took none of these	79	77.5	43.20	5.91	79.42	14.00	42.90	10.75	63.63	10.99
Total	102	100.0	43.28	6.24	80.84	14.83	43.34	11.21	63.39	11.26
t			0.2		1.8		0.7		-0.4	
SPANISH MAJORS										
Took MLA 5-7	399	44.3	43.75	6.79	84.09	12.43	46.96	8.52	54.24	10.46
Took none of these	501	55.7	45.94	6.77	86.05	12.99	49.91	8.74	56.43	10.45
Total	900	100.0	44.97	6.87	85.18	12.79	48.60	8.77	55.46	10.52
t			-4.8**		-2.3*		-5.1**		-3.1**	

** $p<.01$, * $p<.05$

deviations of the NDEA posttest data on the Professional Preparation test; it is possible however, that these smaller standard deviations in the NDEA data were due to the nearer approach of the mean to the ceiling of the test, or to a greater homogeneity of the training experiences of the NDEA Institute teachers after the institute courses.) This finding parallels that for the MLA skills test scores, discussed earlier (see Chapter IV).

2. Estimated National Norms for Teacher Preparation Tests

By procedures similar to those used in Chapter V for the MLA Skills tests, estimated national norms for the MLA Teacher Preparation Tests were developed. They are shown in Tables 10.3 and 10.4. The estimates were made on the assumption, which is manifestly false, that all foreign language majors were enrolled in teacher preparation programs and would be candidates for taking the Teacher Preparation tests. Nevertheless, the estimated norms are the best estimates available for the national population of teacher preparation candidates. Actually, the disparity between the samples used for norming the MLA skills tests and those used for the Teacher Preparation tests affects the interpretation of the skills test scores for teacher preparation candidates. A proper set of norms for all seven MLA tests applicable to teacher preparation candidates would be based solely on samples of those candidates. Use of the estimated norms for MLA skills tests in Chapter V will tend to give a slight handicap to teacher preparation candidates because at least in the case of French and Spanish majors they are somewhat inferior, on the average, to students not enrolled in teacher preparation programs, as shown in Table 10.2. On the other hand, it is perhaps useful to handicap teacher preparation candidates in this way to emphasize their tendency to be inferior in basic language skills.

3. Intercorrelations of MLA Teacher Preparation Tests and Correlations with MLA Skills Tests

Intercorrelations of the Teacher Preparation tests and correlations with Skills tests, for the groups of French, German, Russian, and Spanish majors who took the former tests and also had complete data on the Skills tests are shown in Table 10.5, with comparative data from various documents published by Educational Testing Service (ETS, 1961a, 1961b, 1962a, 1962b) concerning these correlations in NDEA Language Institute samples. The intercorrelations between the Applied Linguistics test and the Civilization and Culture test range from .43 to .58, reflecting possibly the general level of training achieved by the student. The correlations of these two tests with the Professional Preparation test are at about the same level for the Applied Linguistics test (correlations of .50, .51, .19, and .57, the correlation of .19 being for only 23 cases in Russian), and slightly lower for the Culture and Civilization test (correlations of .37, .30, -.12, and .34). One may be fairly well assured that the three Teacher Preparation Tests measure largely different aspects of professional teaching competence.

From the data shown, the Applied Linguistics and the Culture and Civilization tests show moderate correlations with MLA Skills tests, again reflecting, no doubt, the general level of training in the foreign language and associated wisdom achieved by the student. Correlations of the Professional Preparation test with MLA Skills tests tend to be much lower; this test, after all, has little or nothing to do with the knowledge of specific languages since it is all in English and concerns only the principles and procedures whereby any foreign language may be taught.

If account is taken of the differences in the sizes of the samples and the homogeneity of the groups, the results for the CFLT samples are not inconsistent with the data from the NDEA Language Institute teachers.

Table 10.3

Estimated National Norms for MLA Teacher Preparation Tests, College Senior FL Majors (Percentile Ranks of Converted Scores)

FRENCH				GERMAN				RUSSIAN (tentative)			
Applied Linguistics		Culture and Civilization		Applied Linguistics		Culture and Civilization		Applied Linguistics		Culture and Civilization	
Score	%ile Rank	Score	%ile Rank	Score	%ile Rank	Score	%ile Rank	Score	%ile Rank	Score	%ile Rank
69	99.9	67	99.8	69	99.6	70	99.8	64	97.9	71	97.2
68	99.7	66	99.5	68	99.3	69	99.1	63	95.8	70	94.3
67	99.6	65	99.5	67	99.3	68	98.6	62	95.8	69	94.3
66	99.1	64	99.4	66	99.3	67	98.4	61	95.8	68	94.3
65	98.7	63	99.2	65	99.0	66	97.0	60	95.8	67	94.3
64	98.5	62	99.1	64	97.7	65	94.4	59	95.8	66	94.3
63	97.8	61	98.5	63	96.6	64	92.1	58	92.4	65	94.3
62	96.9	60	97.9	62	96.1	63	90.5	57	88.9	64	92.2
61	95.8	59	97.2	61	95.5	62	89.2	56	86.8	63	88.0
60	94.4	58	95.8	60	95.0	61	87.6	55	81.3	62	85.9
59	92.6	57	94.2	59	93.9	60	86.2	54	77.8	61	83.7
58	90.7	56	92.1	58	91.5	59	83.6	53	77.8	60	81.6
57	88.9	55	90.2	57	88.2	58	79.0	52	77.8	59	81.6
56	85.5	54	88.4	56	84.0	57	75.4	51	72.3	58	81.6
55	80.5	53	85.7	55	79.2	56	73.2	50	59.0	57	78.6
54	77.1	52	82.3	54	76.5	55	69.9	49	47.7	56	72.6
53	74.1	51	79.3	53	72.6	54	66.6	48	43.1	55	66.7
52	70.4	50	77.0	52	68.7	53	65.0	47	41.8	54	63.6
51	66.4	49	72.7	51	63.5	52	62.7	46	36.3	53	61.5
50	61.5	48	66.1	50	57.0	51	58.6	45	30.0	52	53.2
49	56.7	47	58.4	49	54.8	50	54.8	44	19.0	51	44.0
48	52.3	46	50.3	48	52.9	49	52.3	43	7.8	50	34.7
47	46.8	45	44.3	47	50.2	48	47.8	42	6.9	49	28.5
46	39.5	44	39.2	46	45.5	47	38.9	41	6.9	48	28.5
45	33.4	43	33.4	45	38.1	46	31.5	40	3.4	47	21.4
44	28.9	42	26.8	44	27.7	45	29.0			46	14.2
43	24.3	41	20.3	43	18.5	44	24.1			35	14.2
42	20.8	40	15.7	42	14.9	43	19.4			44	11.1
41	17.3	39	12.3	41	12.5	42	16.2			43	8.0
40	13.5	38	9.8	40	8.7	41	12.9			42	5.0
39	10.0	37	8.4	39	5.5	40	11.0			41	1.9
38	6.9	36	7.0	38	3.7	39	9.8			40	1.9
37	4.8	35	4.8	37	1.7	38	8.8			39	1.9
36	3.6	34	3.1	36	0.4	37	6.0			38	0.9
35	2.6	33	2.3			36	4.2				
34	1.7	32	1.3			35	2.1				
33	1.1	31	0.8								
32	0.4	30	0.6								
		29	0.4								
		28	0.2								

Table 10.3 (continued)

SPANISH

Applied Linguistics		Culture and Civilization	
Score	%ile Rank	Score	%ile Rank
70	99.9	73	99.9
69	99.8	72	99.0
68	99.8	71	98.1
67	99.8	70	97.5
66	99.4	69	96.4
65	99.0	68	95.4
64	98.9	67	94.5
63	98.6	66	91.5
62	97.9	65	87.8
61	96.9	64	84.6
60	95.8	63	80.2
59	94.4	62	76.5
58	92.7	61	73.3
57	90.9	60	69.1
56	89.6	59	64.9
55	88.5	58	61.2
54	86.3	57	57.5
53	83.2	56	52.7
52	79.6	55	47.1
51	75.8	54	42.2
50	72.2	53	38.6
49	67.3	52	35.5
48	62.3	51	30.5
47	57.4	50	25.6
46	52.6	49	22.2
45	48.1	48	18.9
44	41.5	47	15.7
43	34.1	46	13.0
42	28.7	45	10.9
41	24.5	44	9.1
40	19.2	43	6.9
39	15.4	42	4.7
38	13.3	41	3.5
37	10.7	40	2.4
36	7.4	39	1.1
35	4.6	38	0.2
34	3.4	37	0.2
33	1.9	36	0.1
32	0.4		

Professional Preparation (All Languages)

Score	%ile Rank
80	99.9
79	99.9
78	99.7
77	99.4
76	99.0
75	98.1
74	96.9
73	95.3
72	92.9
71	89.8
70	86.1
69	82.0
68	78.2
67	74.1
66	68.9
65	62.8
64	57.9
63	53.4
62	48.1
61	42.6
60	37.2
59	32.7
58	28.9
57	24.9
56	20.8
55	16.7
54	13.4
53	11.1
52	9.1
51	7.2
50	5.8
49	4.8
48	4.1
47	3.4
46	2.7
45	2.2
44	1.9
43	1.6
42	1.3
41	1.1
40	0.7
39	0.5
38	0.5
37	0.4
36	0.3
35	0.3
34	0.2
33	0.2
32	0.2
31	0.2
30	0.1
29	0.0

Table 10.4

Estimated College Senior Major National Norms
MLA Teacher Preparation Tests
Converted Scores, for Selected Percentile Ranks

Percentile Rank	FRENCH(Est.N=5564)		GERMAN(Est.N=1848)		RUSSIAN*(Est.N=365)		SPANISH(Est.N=4439)	
	Appl. Ling.	Cult.& Civ.	Appl. Ling.	Cult.& Civ.	Appl. Ling.	Cult.& Civ.	Appl. Ling.	Cult.& Civ.
99	66	62	65	69	64	71	65	72
97	62	59	63	66	64	71	61	70
95	60	58	60	65	61	70	59	68
90	58	55	58	63	57	63	56	66
85	56	53	56	60	56	62	54	64
80	55	51	55	58	55	57	52	63
75	53	50	54	57	51	56	51	62
70	52	49	52	55	51	56	50	60
65	51	48	51	53	51	55	49	59
60	50	47	51	51	50	53	48	58
55	49	47	49	50	50	52	46	56
50	48	46	47	48	49	52	45	56
45	47	45	46	48	48	51	45	55
40	46	44	45	47	47	51	44	53
35	45	43	45	46	46	50	43	52
30	44	42	44	45	45	49	42	51
25	43	42	44	44	45	48	41	50
20	42	41	43	43	44	47	40	48
15	41	40	42	42	44	46	39	47
10	39	38	40	39	43	44	37	44
5	37	35	39	36	41	42	35	42
3	35	34	38	35	40	41	34	41
1	33	31	37	35	40	38	33	39
No. cases actually tested	445	460	116	111	18	20	393	404
Mean (est. norms)	48.18	45.97	48.62	50.15	49.16	52.80	46.18	55.42
S.D. (est. norms)	[7.16]	[6.53]	[6.78]	[8.43]	[5.73]	[7.46]	[7.19]	[7.86]

Percentile Rank	Professional Preparation (Est. N=12,362) (All languages)
99	76
97	74
95	73
90	71
85	70
75	67
70	66
65	65
60	64
55	63
50	62
45	61
40	60
35	59
30	58
25	57
20	56
15	55
10	52
5	49
3	46
1	41
No. cases actually tested	982
Mean (est. norms)	61.91
S.D. (est. norms)	[7.44]

*Tentative because of small N

Table 10.5

Intercorrelations of MLA Teacher Preparation Tests and Correlations with MLA Skills Tests, for CFLT Samples as Compared with Results from NDEA Institute Standardization Samples (ETS, 1961a, 1961b, 1962a, 1962b)[1]

FRENCH GROUPS TAKING THE TEACHER PREPARATION TESTS

CFLT N = 435 NDEA N = 1336

					MLA Skills Tests			
		5	6	7	L	S	R	W
Applied Linguistics	5	1.00 [.782]	.50 (.593)	.50 (.542)	.52 (.581)	.54 (.523)	.57 (.629)	.62 (.655)
Civilization and Culture	6	.50 (.593)	1.00 [.741]	.37 (.491)	.53 (.528)	.46 (.460)	.63 (.624)	.55 (.557)
Professional Preparation	7	.50 (.542)	.37 (.491)	1.00 [.668]	.39 (.306)	.30 (.198)	.42 (.323)	.39 (.296)
	Mean	48.87 (50.8)	46.42 (49.8)	62.77 (66.9)	41.98 (42.3)	72.80 (84.1)	46.32 (46.7)	45.54 (46.5)
	S.D.	7.09 (8.6)	6.37 (8.9)	7.11 (6.7)	7.69 (8.7)	9.25 (18.6)	8.44 (10.5)	9.64 (12.5)

GERMAN GROUPS TAKING THE TEACHER PREPARATION TESTS

CFLT N = 128 NDEA N = 297

					MLA Skills Tests			
		5	6	7	L	S	R	W
Applied Linguistics	5	1.00 [.854]	.49 (.671)	.51 (.437)	.35 (.533)	.36 (.561)	.39 (.616)	.49 (.750)
Civilization and Culture	6	.49 (.671)	1.00 [.780]	.30 (.405)	.49 (.502)	.48 (.454)	.54 (.583)	.55 (.605)
Professional Preparation	7	.51 (.437)	.30 (.405)	1.00 [.635]	.17 (.191)	.09 (.141)	.20 (.178)	.19 (.248)
	Mean	49.60 (53.8)	51.34 (54.0)	62.52 (67.3)	44.26 (42.2)	90.03 (84.5)	52.08 (48.7)	54.42 (48.9)
	S.D.	7.20 (9.2)	8.58 (9.2)	7.19 (6.2)	8.94 (8.6)	15.68 (17.7)	10.51 (10.5)	13.56 (15.1)

RUSSIAN GROUPS TAKING THE TEACHER PREPARATION TESTS

CFLT N = 23 NDEA N = 176

					MLA Skills Tests			
		5	6	7	L	S	R	W
Applied Linguistics	5	1.00 [.771]	.58 (.535)	.19 (.444)	.62 (.432)	.53 (.362)	.56 (.394)	.79 (.676)
Civilization and Culture	6	.58 (.535)	1.00 [.750]	-.12 (.372)	.66 (.429)	.48 (.271)	.62 (.331)	.70 (.504)
Professional Preparation	7	.19 (.444)	-.12 (.372)	1.00 [.686]	-.17 (.219)	-.07 (.114)	-.19 (.033)	-.01 (.210)
	Mean	50.22 (48.8)	53.57 (53.1)	60.04 (64.4)	43.57 (41.7)	85.74 (77.5)	44.87 (38.2)	62.57 (54.9)
	S.D.	6.45 (6.7)	8.06 (8.5)	5.26 (7.0)	7.32 (7.3)	16.54 (18.0)	12.60 (11.1)	12.15 (14.3)

Table 10.5, continued

SPANISH GROUPS TAKING THE TEACHER PREPARATION TESTS

CFLT N = 399 NDEA N = 1334

					MLA Skills Tests			
		5	6	7	L	S	R	W
Applied Linguistics	5	1.00 [.775]	.43 (.576)	.57 (.627)	.32 (.502)	.09 (.410)	.31 (.559)	.38 (.623)
Civilization and Culture	6	.43 (.576)	1.00 [.780]	.34 (.548)	.46 (.606)	.28 (.517)	.52 (.673)	.52 (.681)
Professional Preparation	7	.57 (.627)	.34 (.548)	1.00 [.701]	.22 (.383)	.05 (.271)	.16 (.419)	.23 (.457)
	Mean	46.52 (49.0)	55.53 (56.4)	61.17 (66.8)	43.75 (41.6)	84.09 (77.5)	46.96 (44.1)	54.24 (48.7)
	S.D.	7.36 (8.5)	7.74 (9.0)	7.83 (6.4)	6.80 (7.8)	12.44 (18.9)	8.53 (9.5)	10.47 (12.3)

---

[1] NDEA data, given in parentheses, are from posttests, except that self-correlations, given in brackets [ ], are correlations between pre-test and post-test of different forms as an estimate of reliability.

4. Variables associated with performance on the MLA Teacher Preparation Tests

Various tabulations were made in an effort to investigate what variables might be associated with performance on the Teacher Preparation tests. Since foreign language aptitude had been shown to be relevant in explaining some of the variance in the MLA skills test scores, this was investigated also in connection with the Teacher Preparation tests, with results shown in Table 10.6. A considerable number of significant correlations appear in that table, the incidence of significant correlations being largely dependent upon the numbers of cases involved. In French and Spanish, the Applied Linguistics test is best predicted by the MLAT Total Score, but the MLAT Total shows significant correlations also with Culture and Civilization and with Professional Preparation test scores. In German, with a smaller number of cases, the highest correlation with MLAT Total is for the Professional Preparation test. When we examine the correlations for MLAT subtests, it becomes clear that there is a specific relation, in all languages, between MLAT-4 (Words in Sentences, a test of "grammatical sensitivity") and the Applied Linguistics test; the correlations range from .25 (non-significant, for Russian) to .49 (highly significant, for Spanish). These results seem reasonable, although there was no particular reason for expecting them; students with special language aptitude, especially in grammatical sensitivity, are more likely to develop interests in the linguistic facts and principles tested in the Applied Linguistics test, and more generally in the knowledges tested in the Culture and Civilization and the Professional Preparation tests. In addition it must be pointed out that although the MLAT was designed as a special language aptitude test, its scores are to some extent a reflection of general intelligence, and this may be the reason that MLAT scores show some correlation with scores on the Professional Preparation test, which probably is correlated with general scholastic achievement.

The educational level at which the student intended to teach was also a variable associated with performance on the Teacher Preparation tests, as may be seen in Table 10.7. The major trend in this table is the one that we have come to expect from previous findings, namely, that those who include teaching at the college level in their plans do better on the MLA tests than those who aspire only to teaching in elementary or secondary school. This trend was significant for all three Teacher Preparation tests for the French group, but only for the Civilization and Culture test in the case of the Spanish group. (Insufficient numbers of cases were available in German and Russian to justify this type of analysis.)

It will be noted in Table 10.7 that the large majority of those who took the teacher preparation tests were students intending to teach at the high school level. This is due partly to the restriction placed upon taking the Teacher Preparation tests: only students enrolled in programs "expressly oriented toward the preparation of elementary or secondary school language teachers" were supposed to take the tests. It can be inferred, by comparing the figures in Table 10.7 with those in Table 6.7 that many individuals intending to teach only at the college level, and not enrolled in teacher preparation programs, were not allowed to take the Teacher Preparation tests. Therefore, the norms developed here for the Teacher Preparation tests would have to be used with considerable caution if they are applied to individuals intending teaching at the college level.

Time of beginning foreign language study had been shown to be an important variable in explaining variation in performance on MLA Skills tests. The students who took the Teacher Preparation tests were classified according to whether they had begun the study of their major foreign language in grade school, high school, or college, and the means and standard deviations of the scores are shown in Table 10.8. Although there were slight trends similar to those found for the Skills tests, they were weak and only rarely of statistical significance at even the 5% level. (The table gives the exact levels of significance for each comparison.) The results suggest that starting language study early gives only a slight advantage to the student in acquiring those aspects of teacher competence that are tested by the Teacher Preparation tests. It is likely that many of these aspects are learned mainly at the college level, in any event. For example, it is unlikely that the student starting foreign language study at the elementary or secondary school level would be introduced to, or made aware of, many of the specific facts and principles of applied linguistics until he reaches the college level. Likewise, it would appear that

Table 10.6

Correlations of MLAT Subtests with MLA Teacher Preparation Tests

FRENCH (N = 408)

			Correlations		
MLAT Subtest	Mean	S.D.	MLA-5	MLA-6	MLA-7
3 (Spelling Clues)	26.80	9.23	.31**	.27**	.16**
4 (Words in Sentences)	32.48	5.90	.39**	.25**	.25**
5 (Paired Associates)	19.99	4.62	.27**	.25**	.18**
Total	79.26	13.91	.46**	.37**	.27**

GERMAN (N = 102)

			Correlations		
MLAT Subtest	Mean	S.D.	MLA-5	MLA-6	MLA-7
3 (Spelling Clues)	25.56	10.12	.22*	.12	.27**
4 (Words in Sentences)	33.32	6.12	.34**	.04	.29**
5 (Paired Associates)	20.24	5.06	.13	-.04	.29**
Total	79.12	16.91	.29**	.07	.35**

RUSSIAN (N = 17)

			Correlations		
MLAT Subtest	Mean	S.D.	MLA-5	MLA-6	MLA-7
3 (Spelling Clues)	26.94	12.12	-.10	.21	-.32
4 (Words in Sentences)	30.41	7.94	.25	.28	-.02
5 (Paired Associates)	21.24	3.63	.14	-.08	.26
Total	78.59	19.88	.06	.23	-.16

SPANISH (N = 352)

			Correlations		
MLAT Subtest	Mean	S.D.	MLA-5	MLA-6	MLA-7
3 (Spelling Clues)	24.95	10.36	.19**	.05	.10
4 (Words in Sentences)	30.64	7.58	.49**	.28**	.39**
5 (Paired Associates)	18.98	5.61	.22**	.17**	.20**
Total	74.31	16.65	.51**	.33**	.43**

**p < .01; *p < .05 for hypothesis of zero correlation.

Table 10.7

Means and Standard Deviations of MLA Teacher Preparation Test Scores, by Intentions to Teach at Various Levels, for French and Spanish Groups

			Applied Linguistics MLA-5		Culture & Civilization MLA-6		Professional Preparation MLA-7	
	N	%	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ
			FRENCH MAJORS					
Do not intend to teach	18	4.2	49.94	8.82	47.33	8.55	56.78	9.36
Intend to teach at:								
Elem. sch. only	13	3.1	45.54	7.76	41.31	6.11	61.23	6.65
Elem. sch. or high sch.	29	6.8	46.72	6.63	47.10	6.42	62.31	5.66
High sch. only	302	71.2	48.31	6.92	45.55	6.02	62.74	6.97
High sch. or college	50	11.8	52.04	7.44	49.20	6.56	64.38	6.88
College only	12	2.8	48.42	6.92	45.83	6.68	61.92	10.20
Total	424	99.9	48.63	7.23	46.04	6.44	62.58	7.24
F			3.44		4.65		3.16	
P			< .01		< .001		< .01	
			SPANISH MAJORS					
Do not intend to teach	15	4.1	43.33	5.86	53.67	7.26	57.73	7.57
Intend to teach at:								
Elem. sch. only	--							
Elem. sch. or high sch.	20	5.4	47.00	7.26	53.60	7.49	62.80	7.36
High sch. only	275	74.9	46.37	7.09	54.93	7.36	61.37	7.59
High sch. or college	40	10.9	47.50	8.35	57.75	8.13	60.80	8.15
College only	17	4.6	47.12	6.80	60.35	8.78	59.06	10.38
Total	367	99.9	46.44	7.23	55.37	7.67	61.13	7.85
F			0.98		3.52		1.31	
P			> .05		< .01		> .05	

Table 10.8

Mean MLA Teacher Preparation Test Scores, by Language and By Time of Beginning Study of the Major Language

			Teacher Preparation Tests					
			Appl. Ling.		Cult. & Civ.		Prof. Prep.	
	N	%	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ
			FRENCH					
Began in grade school	41	9.5	49.07	5.18	45.68	6.44	60.29	6.73
Began in high school	313	72.3	49.01	7.43	46.44	6.41	63.03	7.23
Began in college	79	18.2	47.20	7.02	44.96	6.39	62.35	7.11
	433	100.0	48.68	7.21	46.10	6.45	62.65	7.22
$F_{2,\ 430}$			2.05		1.76		2.71	
p			.13		.17		.07	
			GERMAN					
Began in high school	58	55.2	50.76	7.16	52.05	8.40	64.31	7.21
Began in college	47	44.8	48.34	7.07	49.45	8.57	61.96	6.14
	105	100.0	49.68	7.25	50.89	8.62	63.26	6.88
$F_{1,\ 103}$			2.94		2.40		3.09	
p			.08		.12		.08	
			SPANISH					
Began in grade school	39	10.1	49.03	7.63	57.44	6.54	62.15	9.50
Began in high school	272	70.1	46.68	7.47	55.23	7.72	61.56	7.72
Began in college	77	19.8	45.32	6.55	55.14	7.91	60.08	7.47
	388	100.0	46.65	7.38	55.43	7.69	61.32	7.90
$F_{2,\ 385}$			3.30		1.48		1.29	
p			.04		.23		.28	

students acquire their erudition in Culture and Civilization mainly at the college level. (This conclusion is supported by the findings to be discussed next.)

To investigate other variables, recourse was made to the techniques of multiple and canonical regression introduced in Chapter VIII. The major variables that had been found to be relevant in explaining the variance of the MLA Skills test scores were included in these analyses, summarized in Table 10.9. The table shows the results of both the multiple regression analyses and the canonical regression analyses.

Let us consider first the multiple regression analyses, in the upper part of the table. Each of the MLA Teacher Preparation tests is predictable to a highly significant extent from information on student background, training history, and language aptitude. The MLAT Total score is generally the most potent predictor, a finding that confirms the interpretations of Table 10.6. (If separate MLAT scores had been included in this analysis, the multiple correlations would doubtless have been significantly higher.) Amount of time spent abroad is the next most potent predictor, especially of scores in the Culture and Civilization test. Number of course hours in college literature courses shows a specific relation to the Culture and Civilization test in all three languages. There is a scattering of other significant beta-weights, but their patterning is not consistent over languages. It is somewhat surprising, however, that number of hours in "miscellaneous courses" shows only small and generally non-significant beta-weights. These courses would have included, presumably, special courses in foreign language pedagogy. The only case in which this variable showed a unique prediction of a Teacher Preparation test was in German, for the Civilization and Culture test. The failure of this variable to predict Teacher Preparation tests cannot be ascribed to homogeneity of the teacher preparation sample with respect to the variable, for indeed, the standard deviations were generally larger than the means (indicating highly positively skewed distributions), and the teacher preparation samples were only slightly more homogeneous in this respect than the complete CFLT samples. The following data will support these observations:

Hours of Miscellaneous Courses

		FRENCH	GERMAN	SPANISH
Teacher Preparation Sample	Mean	3.35	3.58	3.45
	S.D.	4.31	3.68	4.57
	N	397	98	343
Complete CFLT Samples as used in Table 8.2	Mean	3.06	2.60	3.44
	S.D.	4.40	3.57	4.97
	N	1039	389	793

There is perhaps some meaning in the fact that the German teacher preparation sample shows a considerably higher mean number of hours of miscellaneous courses than the complete CFLT sample in German, and also shows a significant beta-weight in predicting Civilization and Culture test scores. Do German teacher preparation programs lay more stress on pedagogical courses than other programs? These results suggest so.

The canonical regression results shown in Table 10.9 amplify and enrich the above interpretations. In German and Spanish, there are two independent canonical correlations that are significant at the 5% level or better, while in French one is highly significant and another one significant at the 10% level. Nevertheless, the relative consistency of the patterns of canonical regression weights across languages tempts one to think that as many as three canonicals are of interest. Sorting them, we re-label them A, B, and C.

Canonical A is a summarization of the three Teacher Preparation tests with generally positive weights: in French, Applied Linguistics and Culture and Civilization are positively weighted, while in German and Spanish, Applied Linguistics and Professional Preparation are emphasized. This combination of the criterion tests is universally well predicted by the language aptitude test (MLAT Total), and seems to represent that portion of the Teacher Preparation test variance that is most closely associated with student traits constituting language aptitude and possibly interest.

Table 10.9

Multiple and Canonical Regression Analyses for Teacher Preparation Tests and Selected Predictor Variables

A. Multiple Regression Analysis (Beta Weights and Multiple Correlations)

	Criterion Variables (MLA Tests)								
	FRENCH MLA			GERMAN MLA			SPANISH MLA		
Predictor Variables	5	6	7	5	6	7	5	6	7
Time Abroad	.11*	.20**	.02	.25*	.29**	.07	.03	.14*	.02
Use of FL at Home	.09	.07	.06	.08	.14	.07	-.08	.12*	-.19**
Semesters Grade Sch.	-.03	-.04	-.10**	.08	-.06	.09	.07	-.02	.06
Semesters High Sch.	.05	.04	.06	.14	.12	.25*	.10	.03	.02
Hrs. College Lang.	.04	.00	.10	.23	.12	.26*	.07	.04	-.01
Hrs. College Lit.	.03	.13**	.05	.04	.30**	-.13	-.04	.27**	.00
Hrs. Misc. Courses	.03	.06	-.01	.08	.19*	-.03	.06	.07	.01
MLAT Total	.45**	.34**	.25**	.38**	.18	.38**	.46*	.29**	.33**
Multiple R:	.49**	.46**	.30**	.50**	.57**	.46**	.51**	.45**	.42**
N	397			98			343		

B. Canonical Regression Analysis (Canonical Regression Coefficients and R's)

	Successive Canonical Regressions, by Language								
	FRENCH			GERMAN			SPANISH		
Original No:	I	II	III	I	II	III	I	II	III
Relabeled:	A	B	C	B	A	C	A	B	C
Predictors (Canonical Regression Weights)									
Time Abroad	.33	.64	.01	.57	.07	-.50	.09	.33	.10
Use of FL at Home	.18	-.07	.08	.24	.04	.23	-.19	.52	-.65
Semesters Grade Sch.	-.07	.19	-.65	-.11	.24	-.23	.12	-.16	-.03
Semesters High Sch.	.10	-.14	.18	.14	.35	.78	.17	-.05	-.42
Hrs. College Lang.	.06	-.40	.45	.17	.43	.18	.11	.04	-.39
Hrs. College Lit.	.17	.41	.51	.58	-.38	-.02	.01	.76	.39
Hrs. Misc. Courses	.09	.25	-.08	.37	-.12	-.09	.12	.12	-.21
MLAT Total	.90	-.37	-.26	.29	.68	-.05	.94	.08	-.21
Canonical R:	.53	.19	.12	.59	.51	.13	.52	.42	.15
p:	< .001	.10	.41	< .001	.01	.95	< .001	< .001	.23
Criteria (Canonical Regression Weights)									
MLA-5 (Appl. Ling.)	.78	-.44	-.66	.29	.50	-.74	.92	-.33	-.65
MLA-6 (Cult. & Civ.)	.62	.77	.28	.91	-.47	.33	.17	.91	.08
MLA-7 (Prof. Prep.)	.03	-.46	.70	-.29	.73	.58	.35	-.25	.76

**p < .01; *p < .05. Significance levels given only for beta-weights, multiple correlations, and canonical correlations.

Canonical B is one which, on the criterion side, features the Civilization and Culture test in contrast with the other tests. In all three languages, it is predicted well by two variables in particular: Amount of Time Abroad, and Hours of College Literature. The former is the more important of these variables in French, the latter is the more important in Spanish, and they are about equally weighted in German. This finding gives confirmation to the expectation that two ways of acquiring superior knowledge of the culture and civilization of a language are (1) to study and travel abroad, and (2) to give protracted attention to the literature of that language.

Canonical C gives emphasis, on the criterion side, to the Professional Preparation Test. It is a very weak source of variance, and there is little consistency in the patterns of canonical regression weights for the predictors. (It should be remembered that despite the small canonical correlations the canonical regression weights are scaled in such a way that the sum of their squares is equal to unity; it is therefore almost inevitable that there will be, by chance, at least one or two coefficients of fairly large magnitude.) We may conclude that the residual variance in the Professional Preparation test remaining after the effects of the other canonicals are taken into account is not reliably predictable.

Chapter XI

SUMMARY AND RECOMMENDATIONS

1. Purpose and Design of the Study

The purpose of this study was to assess the overall levels of foreign language proficiency attained by senior class students "majoring" or concentrating in foreign languages in colleges and universities of the United States, and to identify factors strongly associated with these levels of competence, with the hope that such information might suggest ways in which foreign language teaching for these students could be improved.

The study was necessarily based upon a sample of graduating seniors majoring in foreign languages drawn from the population of such majors and tested at a particular point of time. The target population chosen for this purpose was that of all students who were majoring in French, German, Italian, Russian, and Spanish and graduating in the class of 1965 from colleges and universities in the United States. More precisely, the target population consisted of all students majoring in these languages and graduating from institutions that had been listed in U. S. government publications as offering B. A. programs in foreign languages in the academic year 1962-63. In drawing the population, a two-stage stratified sampling design was employed in which institutions were to be sampled randomly within certain classifications as to size and principal source of financial support, and complete samples of students within these institutions were to be tested. The number of institutions to be sampled within size and institution type classifications was determined by statistical formulas that specified a sample that promised to yield the maximum amount of information with the least amount of error for the planned budget.

As measures of foreign language proficiency, as well as of certain aspects of preparation for teaching (for those students enrolled in programs for preparing teachers of foreign languages), the subtests of Form A of the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students were used: four tests of foreign language skills (Listening, Speaking, Reading, and Writing) and three tests of teacher preparation (Applied Linguistics, Culture and Civilization, and Professional Preparation). As sources of subsidiary information on the students, the "short form" of the Modern Language Aptitude Test and a Questionnaire for Foreign Language Majors were used. The latter inquired in considerable detail into the backgrounds and foreign language learning experiences of the students. A Questionnaire for Foreign Language Department Chairmen sought to gain information on the characteristics of the college or university foreign language departments in which the sample of students had studied.

Parallel to the main study, an ancillary study was performed in an attempt to establish meaningful equivalents for the skill levels measured by the MLA skills tests. For this purpose, the Foreign Service Institute interview procedure was administered to small samples of teachers attending NDEA Language Institutes, the resulting "absolute language proficiency ratings" being equated to scores attained on the MLA skills tests. These samples were chosen in such a way as to represent a wide range of foreign language competence. The FSI-MLA equivalents thus established disclosed that some of the tests, in some languages, fail to measure adequately the upper levels of foreign language competence, i.e., those levels approaching the competence of the educated native speaker.

The main study was conducted with the assistance of Educational Testing Service, which arranged for the administration of the tests and other instruments after the project staff had drawn the sample of institutions. The sampling of institutions was actually done in two phases. From a total of 771 institutions identified as constituting the target population of institutions, a stratified sample of 192 institutions was drawn and invited to participate; it was estimated that these institutions contained a total of approximately 6000 students majoring in the five foreign languages included in the study. Because only 146 of these institutions agreed to participate, a second random drawing of

101 additional institutions was made, making 293 institutions in all that were invited to participate.

2. Outcomes of the Sampling Design

As of April 9, 1965, when the books were closed on the specification of the sample of participating institutions, 203, or 69.3%, of the 293 institutions invited to participate had agreed to do so. There was little difference between public and private institutions in rate of response, but there was a clear tendency for the larger institutions to be more likely to agree to participate. There is no way of knowing to what extent this trend may have biased the sample with respect to the criterion variables.

After the testing, conducted mainly in April and May of 1965, it became evident that student response rate fell considerably short of the planned 100% response rate within participating institutions. That is, although every effort had been made to persuade students at the participating institutions to take the tests, which were given free of charge, by no means every student eligible to take the tests did so. It was not possible to determine the actual student response rate until nearly a year later, when official U. S. Office of Education figures on B. A. degrees earned in 1964-65 became available to the project. At that time, it was determined that there were 5427 students (graduating seniors majoring in one of the five test languages) available to be tested in the 203 institutions that participated in the study. The actual number of students legitimately tested in these institutions (i.e., after 58 underclassmen and 35 graduate students had been excluded) was 2784. The overall student response rate, therefore, was 51.4%. The tested students represented, in all, 23.9% of the 11,633 students in the target population of French, German, Italian, Russian, and Spanish majors present in the class of 1965 in the target population of institutions. Detailed studies of student response rates disclosed that (1) there was no significant difference in overall response rate between men and women, (2) student response rate was negatively correlated with the size of the institution, students in the larger institutions being less likely to volunteer to take the tests than students in the smaller institutions, and (3) the student response rate was higher at private (67.3%) than at public (39.6%) institutions, and this difference was consistent for institutions of different sizes.

The original intention of the study to sample about 50% of the target population of students was thus achieved only about half-way; it was frustrated partly by less-than-perfect institutional response rates and partly by far-from-perfect student response rates. Nevertheless, certain counterbalancing factors produced a sample that was probably fairly representative of the target population. Specifically, the higher institutional response rate for the larger institutions tended to counterbalance the lower student response rates in those institutions. (Further comment on sampling biases will be made in Section 4 of this chapter.) For this reason there is probably little error in using the sample as it stands.

Though there were highly significant differences in response rates over languages, these differences appear to be largely associated with Russian, where the response rate was consistently low (overall: 31.7%) in both public and private institutions.

It was inevitable that the data were not complete for the students actually tested. Only 2587 students had scores on all four MLA skills tests. Only 2534 students took the MLAT. In all, there were 2389 cases complete on the MLA skills tests and the MLAT. Questionnaire data were available on all of these, although it was sparse in a handful of cases.

Only students enrolled in teacher preparation programs were supposed to have taken the teacher preparation sections (5, 6, and 7) of the MLA tests; there were 1116 students who took these additional sections. 980 of these were complete on the four skills tests and the MLAT.

The numbers of "legitimate" cases tested in each of the five languages were as follows: French, 1270; German, 413; Italian, 28; Russian, 105; and Spanish, 968.

Most of the analyses performed in the study utilized only "regular" cases, i.e., students who were apparently native speakers of English who had studied their major foreign language in the course of the normal educational sequence, and had not had special opportunities such as protracted residence abroad (more than one year) or special intensive courses such as are available in the Armed Forces. There were 2604 such "regular" cases.

3. Levels of Proficiency Attained by Students in the Tested Sample

Frequency distributions of scores on the MLA skills tests had a tendency, in many cases, to have a negative skew. These findings, coupled with evidence from the FSI equivalences and from studies of possible curvilinear relations among the tests, suggested that some of the MLA tests--especially the Listening tests--have "ceilings" whereby they do not adequately measure the upper levels of foreign language competence. For example, in terms of the FSI equivalents established for the MLA tests, the Listening test measures only up to S-3+ in French and Russian, S-4 in Spanish, and S-4+ in German. It may be concluded that although the MLA tests are billed as "advanced" tests, in many cases they are not sufficiently difficult to discriminate among really advanced levels of proficiency.

Nevertheless, the scores obtained ranged from scores in the chance range to scores at or near the maximum possible scores. A general characteristic of the tested samples is that they were much poorer in Listening and Speaking skills than they were in Reading and Writing. Typically, the "regular" cases had mean scores in Listening and Speaking that correspond to FSI ratings of S-2 or S-2+, i.e., in the range of "limited working proficiency." In Reading and Writing, however, the tested students tended to have mean scores that correspond approximately to an FSI rating of R-3, "Minimum professional proficiency." There was some variation across languages in the average FSI equivalents attained; French samples appeared to be particularly poor in Listening and Speaking, while German samples did somewhat better in these skills, on the average.

Intercorrelations among the tests were consistent with those previously obtained for samples of NDEA Institute teachers after account is taken of the fact that the CFLT (College Foreign Language Testing) samples are generally more homogeneous in proficiency levels than the NDEA teachers. The intercorrelations were generally high and suggested that a large proportion of the variance of each separate skill test is attributable to a common factor of overall competence in the language tested. The estimated proportions attributable to a common factor were typically between .80 and .90 after correction for unreliability. Nevertheless, it appeared that a small but significant proportion of the variance of each test measured some specific aspect of foreign language skill. The Speaking test generally had the largest proportion of specific variance.

In the main, the regressions of the tests on one another did not depart significantly from linearity. Where they did, the most likely explanation is that one of the tests has a "floor" and/or a "ceiling" such that the discriminative power of the test is inadequate at either the lower or the upper end of the scale of measurement, or both.

4. Estimates of Levels of Foreign Language Proficiency Attained in the National Population

An attempt was made to estimate the amount of bias in the sample with respect to the criterion variables introduced by institution and student response rates that were smaller than desired.

The possible effect of student non-response (i.e., failure of an eligible student to volunteer to take the tests) was investigated by a special study based on data obtained from 15 medium-large institutions that had tested anywhere from 25 to 90 per cent of their eligible students. From transcripts supplied by the registrars, foreign language grade-point-averages were computed for all eligible students. It was found that there was a statistically significant trend for students who participated in the testing to have slightly higher grade-point-averages than students who did not. Through certain statistical procedures, it was estimated, however, that if all students had participated in the testing, the mean MLA skills test scores would have shifted downwards only a very small amount--less than one converted score point, on the average. It was concluded that student non-response made relatively little difference in the national norms--not enough to justify any general correction in the norms.

Implicit in the sample design was the assumption that the mean achievement of students might vary with institution size and institution type (public vs. private). Analysis disclosed that such variation did indeed occur. Students at the larger institutions tended to score higher than students at the smaller institutions, and students at private institutions scored significantly higher, on the average, than students at public institutions. For some language groups (German and Spanish), there was a small but significant interaction between institution size and type; students at the largest private institutions tended to score even higher than one would predict solely on the basis of knowledge of institution size and type.

Further analysis disclosed that some of the differences in mean score between large and small institutions or between private and public institutions could be accounted for by the fact that students who had started the study of their major foreign language early (i.e., in elementary school or high school) were more likely to attend large and/or private institutions. Furthermore, students at the larger, private institutions were more likely to have had travel or study abroad. Time of beginning language study and the amount of time abroad were variables that were shown to be both very strongly associated with the attainment of high foreign language proficiency, particularly the latter variable. Even after the effects of these variables were statistically removed from the differences noted with respect to institution size and type, however, there remained appreciable variation among mean scores of institutions of different sizes and types. Evidence was assembled that suggested that these differences could be explained to a large extent by the fact that the larger and/or private institutions are more likely to have instructors who use the foreign language in the classroom and offer more advanced instruction. It was concluded that the larger institutions, and to some extent the private ones, produce better language students than the smaller and/or public ones, even after student backgrounds are taken into account. They appear to have better facilities for doing so.

Regional differences in mean scores were found. Generally, the mean proficiency test scores were higher in those regions where the languages are more popular. At the same time, students classified by regions appear to differ in their mean overall performance, regardless of language. For example, students in New England institutions are generally high in all languages. Regional differences, it should be pointed out, are probably related in a complex way to other variables that are associated with test scores, for example, time of beginning language study and amount of time abroad; no attempt has been made to study regional differences adjusted for the effects of these other variables.

Estimated national norms were derived by inflating the frequencies in test score distributions for particular institution sizes and types in inverse proportion to the degree to which those institutions were represented in the sample. These norms, interpreted with reference to the Foreign Service Institute proficiency equivalents that had been established, yielded important estimates as to the overall levels of skill likely to be attained by college seniors in the United States majoring in foreign languages. It was striking that according to these estimates, students attain relatively low levels of skill in spoken foreign languages. The median graduate with a foreign language major can speak and comprehend his major language only at about an FSI Speaking rating of "2+," that is, somewhere between a "limited working proficiency" and a "minimum professional proficiency." This is generally true of all the language groups, the French group showing particularly inferior performance. But in skills with the written language (Reading and Writing) the French, German, and Spanish groups attain median scores that correspond approximately to an FSI rating of "R-3," that is, reading and writing skills a "minimum professional proficiency." Russian groups show low average performance even in reading and writing skills, however. It may be concluded that the net showing of college foreign language majors in language skills is far from impressive. On the other hand, there are appreciable numbers of students who exhibit quite superior competence. Furthermore, the college senior populations exhibited clearly higher mean scores in foreign language skills (except Speaking, in some cases) than did samples of foreign language teachers at NDEA Language Institutes, and were more homogeneous than the NDEA Institute teacher groups.

When evaluated against the standards of "teacher qualification" established by Myers and Melton (1964), the college senior samples show up rather poorly. The Myers-Melton standards are very demanding, it would appear. For example, in French, it is estimated that only 16% of the national population qualify as "Superior" in Listening, less than 1% in Speaking, 25% in Reading and about 8% in Writing. But relatively few of the CFLT population would fall into the Unsatisfactory category by the Myers-Melton standards.

Still, whether one uses the FSI ratings or the Myers-Melton ratings as standards, the estimated average level of proficiency of the college senior foreign language major is not as high as might be desired. These individuals are particularly deficient in audiolingual skills. Only about a third to a half are at a satisfactory level even in Reading and Writing skills.

5. Student Factors in Foreign Language Skill Attainment

Studies were made to determine what characteristics of the students might be associated with their foreign language proficiency. All these analyses involved only the students actually tested, or appropriate subgroups thereof.

The majority of foreign language majors are women. In the samples studied, 84% of the French majors were women. The figures for the other languages were: German, 59%; Russian, 62%; and Spanish, 75%. There were no significant and consistent mean differences in tested language proficiency between the sexes (only one difference being statistically significant).

Over the whole CFLT sample, 88.8% were at age 20, 21, or 22; ages of the remainder ranged up to about 50. There was a slight tendency for younger graduates in French and Spanish to have superior scores; this may be a reflection of the fact that more of these had had early exposure to these languages in elementary school programs. There was no strong tendency for older students (23 and over) to obtain inferior scores.

The CFLT groups were on the average quite superior in their language aptitude as tested by the Modern Language Aptitude Test. There were significant differences across languages: the language groups rank in the order Russian, French, German, Spanish. Language aptitude scores show quite small--barely significant--correlations with foreign language proficiency test scores for the complete samples. Further analysis showed, however, that language aptitude is a relatively important variable in language proficiency attainment when certain other influential variables are controlled.

About 65% of the CFLT sample planned to go into the teaching of foreign languages at some level, a slightly higher proportion of women (66%) having such plans than men (61%). By far the largest number of both males and females who plan to teach expect to do so at the secondary school level. Teaching at the college level attracts about 32% of them--more men than women. Teaching at the elementary school level is included in the plans of only 12% of those intending to teach. Of the complete CFLT sample, about 37% hoped to use their foreign language skills in employment opportunities other than teaching. Some type of governmental work attracts the largest proportion of these latter, with business activity running a close second.

Reasons for majoring in foreign languages also include various "fringe benefits," such as the opportunity to use FL knowledge in travel or study abroad, the serving of interests in foreign cultures and literatures, or the promotion of intercultural understanding.

Almost universally, students rated all four language skills (listening, speaking, reading, and writing) as of "great importance." Speaking and listening, however, had higher average ratings than reading and writing. These findings appear to give strong support to contemporary policies in foreign language teaching whereby listening and speaking are accorded emphasis and attention at least equal to that of reading and writing. Students regard reading and writing as "advanced" skills of great importance, particularly for specialized work in foreign languages, but listening and speaking are regarded as being of even greater importance, particularly for teaching at elementary and secondary school levels. French, German, and Spanish majors attach relatively more importance to speaking and listening than do Russian majors. It was shown, further, that the relative amount of importance a student attaches to listening and speaking as contrasted with reading and writing is to some extent related to his own relative competencies in these skills.

Vocational plans are to some extent associated with mean levels of language proficiency attained. Those planning to teach at the secondary school level are significantly inferior, on the average, to the remainder of the group. Those planning to teach at the college level are strikingly superior to others.

Six largely independent ways in which foreign language majors vary are as follows: (1) whether, on the one hand, they plan to teach, at some level, or, on the other, use their foreign language skills in some kind of employment (usually, government or business) other than teaching; (2) if they plan to teach, whether they plan to teach at the elementary or secondary school level; (3) if they plan to teach, whether they plan to teach at the college level; (4) whether they have traveled or studied abroad; (5) whether they are attracted by the intrinsic benefits of foreign language study (e.g. learning about foreign cultures and literatures); and (6) whether they have tended to take other languages besides their major language, especially Latin and/or Greek.

6. Experience and Training Factors in FL Skill Attainment

One of the major variables found to be associated with foreign language skill attainment is the educational level at which the student began study of his major foreign language. About 64% of the sample started in secondary school. However, on the average, students who started in the elementary school were distinctly superior, at graduation from college, to students who started in secondary school, and these in turn were (in most comparisons) distinctly superior to those who started in college. The simplest explanation of this finding is that it is due to differences among the groups in the amount of time they spent in study; there was no evidence that those who started early had any special advantage because of their youth. Starting early did, of course, have the advantage that the student had more time to acquire his skill by the time of college graduation. Time of beginning foreign language study was largely independent of language aptitude; those who started early did not in general have significantly higher language aptitude scores than those who started later.

Further analyses in which the numbers of foreign language courses taken at the various educational levels were studied in relation to levels of skill attained only confirmed the conclusions mentioned above: that the level of skill attained is to a considerable extent a function of the amount of time devoted to foreign language study. However, in some cases there was a suggestion that the taking of college language courses (as opposed to literature courses) by students arriving from high school with insufficient preparation had either no significant effect or even a negative effect on the students.

It was very difficult to obtain reliable information on the kind and quality of instruction that students had had, but there was distinct evidence that a possible favorable influence is the extent to which the teacher uses the foreign language in the classroom and requires the student to use it. Use of a language laboratory did not show any significant effect, nor did the extent to which the teacher was reported to have a native or near-native accent in the language. The major variable in instruction, then, is the amount of time devoted to it, or more precisely, the amount of time during which the student is involved in the use of the foreign language.

Probably the most important variable associated with the attainment of high foreign language proficiency discovered in this study was the amount of time the student had spent abroad in a country where his major language was spoken. Substantial numbers of students in the sample had had such experience. In MLA skills test scores, there were highly significant differences favoring groups who had been abroad, particularly those who had been abroad for as much as a year of study. The effect of time spent abroad on test scores was found to be independent of the effect of time of beginning foreign language study, even though it was also found that those who started language study early were more likely to have been abroad. Amount of time spent abroad was also independent of language aptitude. Although it was impossible to determine whether time spent abroad had a causal influence on language proficiency, there was a clear suggestion that travel and study abroad improve foreign language skill.

Another powerful variable, for the relatively few students for whom it is relevant, is the extent to which the foreign language is used in the student's home. Students who claim that there is "frequent" use of the foreign language at home perform almost as well, on the average, as native speakers, and students who report "occasional" use at home tend to do better than students who report no such use.

There were also indications that various other experiences, such as the taking of informal courses, the use of the foreign language in a job situation, or extensive reading, were associated with higher language proficiency test scores, but it was not clear whether such experiences should be regarded as causes or effects.

The study was not designed to determine whether the study of other modern languages besides the major language had any special effect on proficiency in the major language. Trends in the data suggested that these effects were more often positive than negative, but the positive effects were not necessarily associated with the degree to which the languages themselves were related. It seems likely that students with greater experience in the study of other languages do better simply because they have greater interest in, and aptitude for, foreign language study. There was little evidence to indicate what effect the study of Latin or Ancient Greek might have on proficiency in the student's major modern foreign language. In some comparisons, those who had had Latin in high school tended to have slightly lower Speaking test scores, and those who had had Latin in college tended to have slightly higher Writing test scores.

7. Comparisons of Influences of Student and Experiential Variables

Multiple regression techniques were utilized to assess the independent influences of various student and experiential variables on MLA skills test scores. Multiple correlations using such variables in predicting MLA skills test scores centered around .60 but ranged as high as .73 for some groups. Thus, a very considerable proportion of the variation in MLA skills test scores can be accounted for as being associated with student characteristics and backgrounds. Among all the variables used in this analysis, amount of time spent abroad was generally the most potent predictor, particularly in French, German, and Spanish groups. Only of secondary importance were variables such as language aptitude, amount of formal training, use of the foreign language at home, and certain specific instructional variables such as the extent to which the student had been exposed to teachers who used the foreign language in the classroom.

The superiority of a student in reading and writing, as opposed to his level of skill in listening and speaking, was found to be predictable chiefly from one of the language aptitude subtests that measures "grammatical sensitivity," and also from the extent to which he has not been abroad. (Conversely, the superiority of a student in listening and speaking over his level of skills in reading and writing is predictable from relatively lower scores on the grammatical sensitivity measure and relatively more time spent abroad.)

A student's superiority in productive skills (speaking and writing) as opposed to receptive skills (listening and reading) can also be predicted with some reliability from background variables such as time of beginning foreign language study, amount of time abroad, and use of the foreign language at home. It is interesting that the patterns of relationships differed from language to language; this may mean that they are affected by the ways in which the languages are typically taught--French and Spanish with more emphasis on productive skills and German and Russian with more emphasis on receptive skills.

It was shown that all these patterns of relationships whereby performance on skills tests can be predicted from background variables tend to hold over various classifications of institutions as to size and type. For example, amount of time spent abroad is universally a good predictor regardless of whether the student attended a small or a large institution, or a private or a public institution.

8. Foreign Language Department Characteristics in Relation to Student Achievement

There was significant variation among institutions in the mean scores of their students on the MLA skills tests. To a large extent, these variations were associated with institution size, the larger institutions tending to have the students with the higher mean scores. It was difficult to identify the specific ways in which departments varied, beyond gross characteristics such as department size and the number of languages included within a department. Department chairmen's reports generally failed to agree with student reports on instructional practices. One variable that tended to have some small degree of significant relationship to mean student performance was the degree to which (according to either students' or chairmen's reports) the use of English was avoided in introductory language courses. Other variables, such as whether English was avoided in more advanced courses, whether there was an emphasis on audiolingual skills, how many courses were offered, or how many degree requirements there were, rarely showed any significant relationships with mean student performance. Our analyses of departmental characteristics failed for the most part to disclose the essential differences between "good" and "poor" language departments.

9. Results from the MLA Teacher Preparation Tests Administered to Students Enrolled in Teacher Preparation Programs

Subtests 5, 6, and 7 of the MLA Proficiency Tests presumably measure competencies in certain areas related to foreign language pedagogy: respectively, Applied Linguistics, Culture and Civilization, and Professional Preparation. These subtests were administered only to students enrolled in programs designed specifically to prepare foreign language teachers for the public schools. Of the 2874 students who took any of the MLA tests, 1116 (38.8%) took one or more of the Teacher Preparation tests. It is possible that an abnormally high proportion of teacher preparation candidates volunteered to take the tests in view of the advantages they gained thereby, but no information was developed as to this possibility. In any case, the students who took the teacher preparation tests tended to be somewhat inferior to the other students in MLA skills test performance. This finding agrees with the finding noted elsewhere that those planning to teach in public schools (as contrasted with those expecting to teach at the college level, or not planning to teach at all) tended to have slightly lower mean scores on the skills tests. Nevertheless, they had mean scores on the teacher preparation tests that were typically midway between the pretest and posttest means of teachers enrolled in NDEA Language Institutes, and they were more homogeneous than those teachers.

Correlations among the teacher preparation tests and between these and the MLA skills tests were interpreted as showing that the teacher preparation tests measure quite distinct aspects of competence for teaching and are only incidentally related to the student's overall level of proficiency in pure language skills.

Time of beginning foreign language study showed no dependable relation to performance on the teacher preparation tests; it was concluded that students probably acquire competence in these aspects of teacher preparation principally at the college level.

Performance on the Applied Linguistics test was most closely associated with scores on foreign language aptitude, especially with the scores on the test of grammatical sensitivity. Performance on the Culture and Civilization test was most closely associated with amount of time spent abroad and with number of college literature courses taken. No variables could be identified that showed any dependable relationships with performance on the Professional Preparation test.

10. Recommendations for the Improvement of the Education of Foreign Language Majors

It must be kept in mind that this report concerns a purely descriptive survey of foreign language majors and cannot provide positive evidence as to the causal factors that make for superior foreign language proficiency or teaching competence. Therefore, it cannot make "sure-fire" recommendations concerning the improvement of the training of foreign language majors. It can only point to factors that exhibit strong relationships to proficiency attainment and suggest that some of these factors may have causal influences.

The study found that the foreign language proficiency of the typical graduating senior majoring in a foreign language was considerably poorer, especially in listening and speaking skills, than what might be desired or expected. Nevertheless, it was possible to identify groups of students who had attained satisfactory or even superior levels of proficiency. These groups were likely to have the following characteristics:

(a) They started study of their major foreign language at an early stage, sometimes in the elementary school, but at least as early as secondary school, and then continued their study through the subsequent years. That is, they devoted a large amount of time to formal study of the foreign language.

(b) They took many courses in literature at the college level.

(c) At either the high school or the college level, or both, they had teachers who made much use of the foreign language in conducting classes and required their students to use it.

(d) They had traveled and studied abroad in a country where their major foreign language was regularly used as the medium of communication. (The data suggested, indeed, that travel and study abroad was the most important single influence in enhancing foreign language proficiency, and that a year's travel and study abroad had much more value than merely a summer tour. Further, it had a stronger influence than starting study of the foreign language at an early age.)

(e) They had various other experiences that encouraged use of the foreign language, such as the taking of informal courses outside regular school work, being employed in a job situation requiring use of the language, or voluntarily devoting much time to the reading of books in the foreign language.

If it is desired to improve the average proficiency of foreign language majors, it would seem advisable to take steps such that more students will have the advantages apparently enjoyed by those who exhibited high levels of skill. That is, more young people should be encouraged to begin study of a foreign language at an early stage of their education and to continue its study through the college years, not only through formal courses but also, and most emphatically, through such experiences as travel and study abroad. At the same time, the teaching of foreign languages should be improved, by strengthening high school and college foreign language departments (especially those at the smaller institutions) so that more courses will be offered and so that more use will be made of foreign languages in the classroom. Since foreign language aptitude proved to be of some significance in the attainment of certain foreign language skills, consideration should be given to the greater use of foreign language aptitude tests in guiding students into foreign language study, especially if they plan to become foreign language teachers.

On the basis of findings of the study, one can make no special recommendations concerning teacher preparation programs other than those already made. Obviously, prospective foreign language teachers must have high levels of proficiency, and all the recommendations we have made concerning the improvement of foreign language skills apply with equal or greater force to teaching candidates. Superior foreign language aptitude seems to be particularly relevant to the attainment of proficiency in applied linguistics, and both courses in foreign language literature, and travel and study abroad, are especially pertinent to the attainment of the requisite knowledge of foreign cultures and civilizations. Any methods that might be found to enhance foreign language aptitude, particularly in the student's awareness of the grammatical aspects of language in general, might be useful in the preparation of teaching candidates. If the skills and knowledges measured by the MLA Professional Preparation test contribute to teaching success, efforts should be made to improve the teaching of those skills and knowledges; the study was unable to find any special factors in student backgrounds associated with performance on this test.

11. Recommendations for Further Studies of Foreign Language Education

The MLA Proficiency Tests

The study disclosed that many of the MLA Proficiency Tests failed to have enough range of difficulty to measure truly advanced levels of skill. The Listening tests were especially defective in this regard, even though the research data suggested that they were the best overall measures of foreign language proficiency. There were suggestions in the data that the Speaking tests are insufficiently reliable and too subject to variations in rating standards and/or the examinee's set or attitude toward the test to serve as really accurate measures of speaking fluency. In some cases the Reading and Writing tests were not of sufficient difficulty. It would be desirable to improve the MLA Proficiency tests both for general use and for use in any further research studies similar to the one just now completed.

It is definitely recommended that further studies be done of the equivalence of the MLA Proficiency Tests to FSI-type "absolute language proficiency ratings," in as much as the equivalences established in the present study were based on samples that were not as large as might be desired. Samples of at least 100 cases, rectangularly distributed with respect to levels of FSI ratings, should be employed, and it might be advisable to delay these equivalence studies until the measurement characteristics of the MLA tests have been improved.

Problems of Sampling and Test Administration

Further studies along the lines of the present one, whether in foreign language education or in other curriculum areas, should give greater attention to problems of sampling, particularly with a view to obtaining better institutional and student response rates. It is probable that such studies should utilize:

(a) somewhat smaller and more carefully selected samples;

(b) a longer "lead time" in order to prepare institutions to participate more effectively;

(c) better publicity for the study.

The two-stage stratified optimal sampling design used in the present study could be again employed, but its success depends upon securing more adequate institutional and student response rates. A study in which only 25% of the population of foreign language majors would be tested would be entirely adequate provided that response rates approached those planned in the design. Those conducting the study could work through professional associations (such as the American Council on Education, the Modern Language Association, and the new American Council of Foreign Language Teachers) in order to secure the desired institutional response rates, and foreign language departments should have more adequate planning time, circulate better publicity about the purposes of the study, and work more closely with students in order to insure that nearly every eligible student takes the tests and other research instruments. It would be advisable to plan to give the tests somewhat earlier in the senior year than was done in the present study--perhaps as early as October or November; it is unlikely that additional learning after that time would significantly affect the test results.

Questionnaire Design

In the course of analyzing the data from the questionnaires completed by students and by department chairmen, it became evident that certain questions were unnecessarily ambiguous, that some questions were probably of little importance, and that insufficient information was obtained on certain matters. A careful perusal of this report will suggest many ways in which the questionnaire design could be improved in future studies.

One egregious omission from the student questionnaire was that of questions specifically pointed towards finding out whether the student was a native speaker of English or of some other language--perhaps the language in which he was majoring, and if he was a native speaker of the major language, whether he had maintained an approximation to native speaker proficiency throughout his life.

Also, the questionnaire should include items that would disclose more precisely the sequence in which the student had various foreign language experiences. If possible, it should ask not only exactly how early the student began his major foreign language, but also how early he may have begun the study of other languages, and how long he had studied them. It should ask how much language study the student had had and how proficient the student estimated himself to have been prior to periods of travel and study abroad. The answers to questions about the history of the student's formal training in the major language should be made more easily codable than was the case in the present study. Questions on the student's plans as to teaching and other uses of his foreign language should be more clearly separated from questions about "reasons" for majoring in a foreign language.

Experimental Design

Although further descriptive, evaluative studies like the present one would be useful, particularly for the purpose of long-range cross-sectional comparisons over a period of years, many questions concerning the education and training of foreign language students would have to be approached through experimental or longitudinal studies in which the effects of various types of learning experiences would be assessed by comparing pretest and posttest performances. For example, we need to know more precisely what the effects of travel and study abroad are, and what level of preparation the student should have for him to obtain maximal advantage from such study abroad. Better studies of the effects of early language training are also needed. Such information could materially aid in the better design of foreign language educational programs.

REFERENCES

Axelrod, Joseph (Ed.) *The education of the modern foreign language teacher for American schools: an analysis of ends and means for teacher-preparation programs in modern foreign languages, based on a study of NDEA Foreign Language Institutes*. New York: Modern Language Association of America, 1966.

Buros, O. K. *The sixth mental measurements yearbook*. Highland Park, N. J.: Gryphon Press, 1965.

Carroll, John B. The prediction of success in intensive foreign language training. Chapter 4, pp. 87-136 in Robert Glaser (Ed.) *Training research and education*. Pittsburgh: University of Pittsburgh Press, 1962; New York: John Wiley, 1965.

Carroll, John B. A model of school learning. *Teachers College Record*, 1963, *64*, 723-733. (a)

Carroll, John B. Research on teaching foreign languages. Chapter 21, pp. 1060-1100 in N. L. Gage (Ed.) *Handbook of research on teaching*. Chicago: Rand McNally, 1963. (b)

Carroll, John B. *Final report: the foreign language proficiency of language majors near graduation from college. Phase I (pilot study)*. Cambridge, Mass.: Laboratory for Research in Instruction, Graduate School of Education, Harvard University, January 31, 1965.

Carroll, John B. Research in foreign language teaching: the last five years. Pp. 12-42 in Robert G. Mead, Jr. (Ed.) *Language teaching: broader contexts. Northeast Conference on the Teaching of Foreign Languages. Reports of the Working Committees*. New York: MLA Materials Center, 1966.

Carroll, John B.; Clark, John L. D.; Goddu, Roland J. B.; Edwards, Thomas M.; and Handrick, Fannie A. *A parametric study of language training in the Peace Corps*. Cambridge, Mass.: Laboratory for Research in Instruction, Graduate School of Education, Harvard University, 1966.

Carroll, John B. & Sapon, Stanley M. *Modern language aptitude test, Form A*. Test booklet, practice exercise sheet, answer sheet, answer keys, manual, tape. New York: The Psychological Corporation, 1958, 1959.

Carton, Aaron S. & Carroll, John B. *The 1959 Summer Russian Language Learning Program: final report of testing project*. Cambridge, Mass.: Graduate School of Education, Harvard University, July 1960. (Prepared for the Inter-University Committee on Travel Grants)

Cochran, William G. *Sampling techniques*. (2nd. ed.) New York: John Wiley, 1963.

Cooley, William W. & Lohnes, Paul R. *Multivariate procedures for the behavioral sciences*. New York: John Wiley, 1962.

Ebel, R. L. Estimation of the reliability of multiple ratings. *Psychometrika*, 1951, *16*, 407-424.

Educational Testing Service. *Test Analysis, Modern Language Association Proficiency Tests, French, JML1, JML2*. Princeton, N. J.: ETS, December 1961. (SR-61-47). (a)

Educational Testing Service. *Test Analysis, Modern Language Association Proficiency Tests, German, JML1, JML2*. Princeton, N. J.: ETS, December 1961. (SR-61-50). (b)

Educational Testing Service. [*Test Analysis*], *Modern Language Association Proficiency Tests, Russian, JML1, JML2*. Princeton, N. J.: ETS, January 1962. (SR-62-3). (a)

Educational Testing Service. Test Analysis, Modern Language Association Proficiency Tests, Spanish, JML1, JML2. Princeton, N. J.: ETS, January 1962. (SR-62-4). (b)

Educational Testing Service. Test Analysis, Modern Language Association Proficiency Tests, Italian, JML1, JML2. Princeton, N. J.: ETS, February 1962. (SR-62-6). (c)

Educational Testing Service. Test Analysis, Modern Language Association Proficiency Tests, Professional Preparation, JML1, JML2. Princeton, N. J.: ETS, February 1962. (SR-62-7). (d)

Educational Testing Service. A description of the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students. Princeton, N. J.: ETS, 1964. (a)

Educational Testing Service. Percentile norms, MLA Teacher Testing Program, MLA Foreign Language Proficiency Tests for Teachers and Advanced Students. Princeton, N. J.: ETS, 1964. (b)

Gulliksen, Harold. Theory of mental tests. New York: John Wiley, 1950.

Handrick, Fannie A. Characteristics of modern foreign language departments and their relation to student achievement. Unpublished Ed. D. Thesis, Graduate School of Education, Harvard University, 1967.

Hansen, Morris H.; Hurwitz, William N.; and Madow, William G. Sample survey methods and theory. Vol. I: Methods and applications. New York, John Wiley, 1953.

Harman, Harry H. Modern factor analysis. Chicago: University of Chicago Press, 1960.

Hays, William L. Statistics for psychologists. New York: Holt, Rinehart and Winston, 1963.

Jones, Kenneth J. The multivariate statistical analyzer. Cambridge, Mass.: [Author], 1964.

MacAllister, Archibald T. (Ed.) The preparation of college teachers of modern foreign languages: a conference report. Publications of the Modern Language Association of America, 1964, 79 (2), 29-43.

Mason, Paul L. & Rice, Mabel C. Earned degrees conferred 1964-65, bachelor's and higher degrees. Washington, D. C.: U. S. Office of Education, Department of Health, Education and Welfare, 1967. (Publication OE-54013-65)

McNemar, Quinn. Psychological statistics. (3rd ed.) New York: John Wiley, 1962.

Myers, Charles T., and Melton, Richard S. A study of the relationship between scores on the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students and ratings of teacher competence. Princeton, N. J.: Educational Testing Service, 1964. (TDR-64-4)

Paquette, F. André (Compiler). Guidelines for teacher education programs in modern foreign languages--an exposition. Modern Language Journal, 1966, 50 (6), 323-425.

Rice, Frank A. The Foreign Service Institute tests language proficiency. Linguistic Reporter, 1959, 1 (2), 4.

Starr, Wilmarth. Proficiency tests in modern foreign languages. Publications of the Modern Language Association, 1961, 76 (2), 7-11. (a)

Starr, Wilmarth. Competency first: new tests in foreign languages. Pp. 97-110 in John B. Carroll (ed.) Proceedings of the Invitational Conference on Testing Problems, 1960. Princeton, N. J.: Educational Testing Service, 1961. (b)

Educational Testing Service. Test Analysis, Modern Language Association Proficiency Tests, Spanish, JML1, JML2. Princeton, N. J.: ETS, January 1962. (SR-62-4). (b)

Educational Testing Service. Test Analysis, Modern Language Association Proficiency Tests, Italian, JML1, JML2. Princeton, N. J.: ETS, February 1962. (SR-62-6). (c)

Educational Testing Service. Test Analysis, Modern Language Association Proficiency Tests, Professional Preparation, JML1, JML2. Princeton, N. J.: ETS, February 1962. (SR-62-7). (d)

Educational Testing Service. A description of the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students. Princeton, N. J.: ETS, 1964. (a)

Educational Testing Service. Percentile norms, MLA Teacher Testing Program, MLA Foreign Language Proficiency Tests for Teachers and Advanced Students. Princeton, N. J.: ETS, 1964. (b)

Gulliksen, Harold. Theory of mental tests. New York: John Wiley, 1950.

Handrick, Fannie A. Characteristics of modern foreign language departments and their relation to student achievement. Unpublished Ed. D. Thesis, Graduate School of Education, Harvard University, 1967.

Hansen, Morris H.; Hurwitz, William N.; and Madow, William G. Sample survey methods and theory. Vol. I: Methods and applications. New York, John Wiley, 1953.

Harman, Harry H. Modern factor analysis. Chicago: University of Chicago Press, 1960.

Hays, William L. Statistics for psychologists. New York: Holt, Rinehart and Winston, 1963.

Jones, Kenneth J. The multivariate statistical analyzer. Cambridge, Mass.: [Author], 1964.

MacAllister, Archibald T. (Ed.) The preparation of college teachers of modern foreign languages: a conference report. Publications of the Modern Language Association of America, 1964, 79 (2), 29-43.

Mason, Paul L. & Rice, Mabel C. Earned degrees conferred 1964-65, bachelor's and higher degrees. Washington, D. C.: U. S. Office of Education, Department of Health, Education and Welfare, 1967. (Publication OE-54013-65)

McNemar, Quinn. Psychological statistics. (3rd ed.) New York: John Wiley, 1962.

Myers, Charles T., and Melton, Richard S. A study of the relationship between scores on the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students and ratings of teacher competence. Princeton, N. J.: Educational Testing Service, 1964. (TDR-64-4)

Paquette, F. André (Compiler). Guidelines for teacher education programs in modern foreign languages--an exposition. Modern Language Journal, 1966, 50 (6), 323-425.

Rice, Frank A. The Foreign Service Institute tests language proficiency. Linguistic Reporter, 1959, 1 (2), 4.

Starr, Wilmarth. Proficiency tests in modern foreign languages. Publications of the Modern Language Association, 1961, 76 (2), 7-11. (a)

Starr, Wilmarth. Competency first: new tests in foreign languages. Pp. 97-110 in John B. Carroll (ed.) Proceedings of the Invitational Conference on Testing Problems, 1960. Princeton, N. J.: Educational Testing Service, 1961. (b)

Starr, Wilmarth. MLA Foreign Language Proficiency Tests for Teachers and Advanced Students. Publications of the Modern Language Association, 1962, 77 (4, Part 2), 31-42.

Tollinger, Suzanne & Paquette, F. André. The MLA Foreign Language Proficiency Tests for Teachers and Advanced Students: a professional evaluation and recommendations for test development. [New York]: Modern Language Association of America, 1966. [Document for restricted use.]

Tolliver, Wayne E. Earned degrees conferred, 1960-61, bachelor's and higher degrees. Washington, D. C.: U. S. Department of Health, Education and Welfare, 1963. (Office of Education Circular No. 721, OE-54013-61). (a)

Tolliver, Wayne E. Earned degrees conferred, 1961-62, bachelor's and higher degrees. Washington, D. C.: U. S. Department of Health, Education and Welfare, 1963. (Office of Education Circular No. 719, OE-54013-62). (b)

Wright, Patricia. Earned degrees conferred 1962-63, bachelor's and higher degrees. Washington, D. C.: U. S. Office of Education, Department of Health, Education and Welfare, 1965. (Publication OE-54013-63, Circular No. 777)

Wright, Patricia. Earned degrees conferred 1963-64, bachelor's and higher degrees. Washington, D. C.: U. S. Office of Education, Department of Health, Education and Welfare, 1966. (Publication OE-54013-64, Miscellaneous No. 54)

Appendix A

RESULTS FROM THE NEW YORK STATE PILOT STUDY

Phase I of this study consisted of a pilot investigation in which an attempt was made to administer the MLA Proficiency Tests, the short form of the MLAT, and a preliminary form of the student questionnaire to a 100% sample of the foreign language majors in French, German, Italian, Russian, and Spanish graduating in the class of 1964 from all institutions in the state of New York offering bachelor's degree programs in those languages. A preliminary form of the Questionnaire for Chairmen of Foreign Language Departments was also tried out. The purpose was primarily to develop experience that would be useful in formulating procedures for the national study, and secondarily to obtain normative data for the students in the New York state sample. A preliminary report of the Phase I study (Carroll, 1965) presented results on response rates, distributions of test scores, and recommendations for the conduct of Phase II of the study (i.e., the national study with which the present report is concerned), and promised that detailed analyses would be included in the final report on Phase II. This appendix presents a recapitulation of the findings previously reported and detailed analyses of Phase I data that are believed to be of interest and pertinence.

Sample design and response rates

The statistical universe was defined as consisting of all students enrolled in the senior class graduating in 1964 in every institution of higher learning in the State of New York listed by the New York State Education Department as having in the academic year 1963-64 a bachelor's degree program in any of the five languages to be tested (French, German, Italian, Russian, and Spanish). (In addition, the Phase I study attempted to utilize the population consisting of all senior-class students, Class of 1964, enrolled as concentrating in any one of these languages in Harvard and Radcliffe Colleges. Only 4 out of 51 students in this latter population volunteered to take the tests, and because of the low student response rate the results will be ignored.)

The New York State Education Department identified 70 institutions in the population. Of these, 44 (63.0%) agreed to participate in the study, 5 (7.1%) had no students to be tested, 1 (1.4%) initially indicated willingness to participate but had to withdraw for reasons beyond its control, and 20 (28.5%) declined to participate or failed to reply. Including the institutions that had no students to test, 49 institutions may be said to have participated, representing an overall response rate of 70.1%. This figure, incidentally, is very close to that reported in Chapter III for all institutions in the national study, 69.3%, but the national figure does not include institutions that had no students to test. (In the national study, the response rate of New York state institutions was 77.4%, a figure that may reflect the fact that New York State has had a keen interest in the MLA teacher certification program and also the fact that in the national study, the New York state institutions not only had acquired experience in participating in our studies but also had somewhat more time for planning their participation, since they received invitations to participate earlier in the year than they did in the pilot study.)

For purposes of comparison with the national study, Table A.1 has been constructed to parallel Table 3.1; it includes only institutions listed in U. S. Office of Education tabulations for 1963-64. On this basis, the institutional response rate for the pilot study was only 62.1% as compared to 69.3% for the national study. The difference may be attributable to the fact that the national study provided institutions more lead time for planning their participation than did the pilot study, in which invitations to participate were not sent out until late in March (March 23, 1964). Otherwise, the data for the New York pilot study are not inconsistent with those for the national study. There is the same tendency for larger institutions to be more likely to participate, although this tendency is not statistically significant for the New York data (Chi-square = 6.5, 4 d. f., $p < .05$).

Table A.1

Institutional Response Rates, New York Pilot Study (1963-64), with Comparative Data from the National Study

Stratum	Size[1]	Solicited Final Sample (= Population) N[2]	Participating N[3]	Participating % of Sample (Response Rates)	Response Rates, National Study
			PUBLIC		
5	40+	5	2	40.0	80.0
4	20-39	2	1	50.0	68.4
3	10-19	0	0	--	70.8
2	5-9	0	0	--	55.0
1	0-4	2	2	100.0	64.7
Total		9	5	55.6	68.6
			PRIVATE		
5	40+	5	4	80.0	90.9
4	20-39	11	9	81.8	86.7
3	10-19	16	12	75.0	69.6
2	5-9	13	8	61.5	70.0
1	0-4	12	3	25.0	56.9
Total		57	36	63.2	69.7
			TOTAL		
5	40+	10	6	60.0	83.3
4	20-39	13	10	76.9	79.6
3	10-19	16	12	75.0	70.0
2	5-9	13	8	61.5	65.1
1	0-4	14	5	35.7	58.7
Total		66	41	62.1	69.3

---

[1] "Size" of institution is defined in terms of the total number of FL earned degrees reported at the institution, according to U. S. Office of Education tabulations.

[2] Includes only institutions listed in U. S. Office of Education tabulations.

[3] Does not include 3 institutions that participated in the pilot study but were not listed in U. S. Office of Education tabulations.

The overall student response rate within institutions was given as 76.2%; this figure is based on the fact that the total number of students tested in the 44 participating institutions was 667 out of 875 individuals reported by department chairmen as being available for testing. The 667 tested students comprised 46.1% of the 1443 students estimated to be present in the statistical universe in New York State. In the national study, the overall student response rate (Table 3.3) was given as 51.3%, and the tested individuals comprised only 24.0% of the statistical universe. The student response rate in the New York pilot study was considerably higher than in the national study. The per cent of the statistical universe tested in the pilot study was also better than in the national study, but this is partly due to the fact that the national study employed a two-stage sampling design in which only a limited number of institutions were invited to participate.

The figures on student response rates given in the preceding paragraph were based on department chairmen's reports of the numbers of students available to be tested. When official U. S. Office of Education figures on numbers of earned bachelor's degrees became available, it was possible to revise the estimates of student response rates. However, U. S. Office of Education tabulations did not cover exactly the same institutions identified by the New York State Education Department. Furthermore, there was considerable ambiguity as to whether the students actually tested were all actually graduating seniors. Table A.2 has been prepared to parallel Table 3.3 for the national study, but it was difficult to adjust the figures so as to be exactly comparable. (That is, in a few cases it would appear that the student response rates were greater than 100%, apparently due to the fact that non-seniors were tested.) Table A.2 nevertheless shows similar trends to those exhibited by Table 3.3: student response rate was higher at private institutions, and it tended to be higher at the smaller institutions. These trends thus appear to be inevitable characteristics of sampling survey studies of this type.

Levels of attainment on MLA tests, tested students

Norms based on the New York state pilot study were presented in the Phase I report but may be regarded as superseded by the national norms presented in the present report. The performance of the New York pilot study groups is summarized, however, in Table A.3, in which the means and standard deviations of all cases complete on the MLA skills test scores, and of all teacher preparation cases complete on all MLA tests, are given. (Four cases from the Harvard-Radcliffe testing were also included here.) The table also shows comparative data from NDEA Language Institute samples as published by Educational Testing Service (Appendix E-3). We may quote the following generalizations about Table A.3 from the Phase I report:

> "1. In French, pilot study cases were superior to both pretest and posttest NDEA norms in all four skills except speaking, where they only barely surpassed the pretest mean.
>
> 2. In German, pilot study cases were very clearly superior to even the NDEA posttest norms, especially in reading and writing.
>
> 3. The few Italian cases (N = 10) were for all skills somewhat below even the NDEA pretest standards.
>
> 4. In Russian, pilot study cases were superior to both pretest and posttest NDEA norms in reading and writing; in listening and speaking they were between the two sets of norms.
>
> 5. In Spanish, the pattern of means for the pilot study cases was similar to that in French; they were superior to both pretest and posttest NDEA norms in all four skills except in speaking, where they barely surpassed the pretest mean.
>
> 6. The means for teacher preparation cases were in general very similar to those for all cases. Thus, prior to exact statistical tests, we judge the teacher preparation cases to be a representative sample of all cases tested."

Table A.2

Student Response Rates, by Strata and Institution Type, New York Pilot Study
(For comparative data on the national study, see Table 3.3)

Stratum	Size[1]	Population (1963-64)[2]	FL Majors in Participating Institutions	% of Population	Tested Students[3]	% of Sample (Response Rates)	% of Population
			PUBLIC INSTITUTIONS				
5	40+	451	170	37.7	78	45.9	17.3
4	20-39	48	20	41.7	22	110.0	45.8
3	10-19	--	--	--	--	--	--
2	5-9	--	--	--	--	--	--
1	0-4	5	5	100.0	11	220.0	220.0
Total		504	195	38.7	111	56.9	22.0
			PRIVATE INSTITUTIONS				
5	40+	313	211	67.4	119	56.4	38.0
4	20-39	299	235	78.6	156	66.4	52.2
3	10-19	211	145	68.7	127	87.6	60.2
2	5-9	83	51	61.4	68	133.3	81.9
1	0-4	31	10	32.3	13	130.0	41.9
Total		937	652	69.6	483	74.1	51.5
			TOTAL				
5	40+	764	381	49.9	197	51.7	25.8
4	20-39	347	255	73.5	178	69.8	51.3
3	10-19	211	145	68.7	127	87.6	60.2
2	5-9	83	51	61.4	68	133.0	81.9
1	0-4	36	15	41.7	24	160.0	66.7
Total		1441	847	58.8	594	70.1	41.2

[1] "Size" in terms of number of majors in all languages in 1962-63.

[2] Restricted to institutions in U. S. Office of Education tabulations.

[3] Includes a few non-seniors.

Table A.3

Comparison of Pilot Study Samples with NDEA Institute Norms of Subtests of the MLA FL Proficiency Tests

	NDEA Norms						Pilot Study					
	Pretest			Posttest			All Cases			Teacher Prep. Cases		
	N	$\overline{X}$	σ	N	$\overline{X}$	σ	N	$\overline{X}$	σ	N	$\overline{X}$	σ
FRENCH												
Listening	4720	38.1	8.8	4770	42.8	8.6	330	44.0	7.6	159	43.5	7.8
Speaking	4494	72.1	19.3	4577	81.0	17.8	330	73.8	12.5	159	72.9	12.8
Reading	4722	43.3	10.8	4770	45.7	10.5	330	49.3	9.0	159	48.9	8.9
Writing	4723	43.0	12.9	4769	45.7	12.6	330	47.6	10.1	159	47.4	9.9
App. Ling.	4641	44.9	8.1	4612	51.1	8.9	--	--	--	159	49.0	7.6
Civ. & Cult.	4640	43.6	8.5	4611	49.3	8.7	--	--	--	159	48.6	7.3
Prof. Prep.*	10771	59.5	8.2	10909	67.0	6.5	--	--	--	159	61.0	7.6
GERMAN												
Listening	1102	38.4	9.1	1256	42.1	9.1	66	45.0	8.0	21	44.3	7.0
Speaking	1086	79.5	17.7	1232	85.3	19.4	66	91.8	16.4	21	87.1	13.1
Reading	1103	44.4	11.9	1256	48.1	10.9	66	53.3	10.1	21	51.6	9.2
Writing	1103	45.8	16.5	1258	48.3	14.6	66	55.4	14.8	21	55.0	13.4
App. Ling.	1102	46.8	9.1	1179	53.0	9.0	--	--	--	21	48.4	10.3
Civ. & Cult.	1101	46.7	10.0	1178	53.4	9.5	--	--	--	21	52.9	8.1
Prof. Prep.*	10771	59.5	8.2	10909	67.0	6.5	--	--	--	21	59.2	8.8
ITALIAN												
Listening	64	40.0	6.0	64	40.5	5.9	10	38.8	7.2	6	37.7	5.2
Speaking	62	88.2	15.4	62	98.0	14.8	10	80.7	16.5	6	81.8	15.8
Reading	64	45.2	10.3	64	47.5	10.7	10	39.8	9.7	6	38.0	5.0
Writing	64	51.1	14.4	64	55.3	14.7	10	46.1	14.4	6	44.0	7.6
App. Ling.	64	46.1	6.5	64	50.7	7.8	--	--	--	6	46.2	5.7
Civ. & Cult.	64	49.8	8.0	64	54.5	8.5	--	--	--	6	44.3	4.3
Prof. Prep.*	10771	59.5	8.2	10909	67.0	6.5	--	--	--	6	53.7	7.5
RUSSIAN												
Listening	597	38.4	7.0	587	43.3	7.5	42	42.6	6.4	6	43.8	5.1
Speaking	584	73.0	17.7	570	87.1	17.0	42	79.2	16.6	6	79.2	12.1
Reading	556	35.5	10.7	545	38.2	10.8	42	43.4	13.5	6	44.8	14.5
Writing	555	48.6	16.5	545	53.4	14.8	42	59.8	11.7	6	66.3	6.4
App. Ling.	515	42.8	6.8	505	47.6	6.7	--	--	--	6	46.2	5.1
Civ. & Cult.	556	47.2	9.2	545	51.9	8.4	--	--	--	6	50.3	13.4
Prof. Prep.*	10771	59.5	8.2	10909	67.0	6.5	--	--	--	6	55.0	8.7
SPANISH												
Listening	4394	39.3	8.0	4400	41.8	7.6	181	43.5	6.9	98	43.9	6.4
Speaking	4271	72.1	20.4	4316	78.2	17.5	181	74.9	13.8	98	76.0	11.5
Reading	4446	42.2	10.2	4544	44.6	9.5	181	46.7	9.1	98	47.2	8.0
Writing	4444	46.2	13.7	4547	50.2	12.8	181	53.0	11.1	98	54.1	9.9
App. Ling.	4393	43.3	7.8	4494	49.4	8.5	--	--	--	98	45.3	6.6
Civ. & Cult.	4342	49.6	9.7	4443	56.2	8.9	--	--	--	98	55.7	8.5
Prof. Prep.*	10771	59.5	8.2	10909	67.0	6.5	--	--	--	98	61.5	6.8

*The NDEA Institute norms for the professional preparation test, which is identical for all languages, are not given separately by languages.

Concerning the teacher preparation tests, the Phase I report noted:

> "Means for the teacher preparation tests for the teacher preparation cases in the pilot study were in general between the corresponding pretest and posttest means in the NDEA norms. The means for Applied Linguistics and Civilization and Culture tests were generally higher than those for the Professional Preparation test, and in the case of the German, Italian, and Russian pilot study samples, they were somewhat below even the NDEA pretest norms."

It is now possible to compare the New York pilot study data with those assembled in the national study. Table A.4 has been prepared to parallel Table 4.1. By comparing these two tables, it may be seen that the New York data are in general quite comparable to those from the national study. Critical ratio tests of mean differences disclosed no significant differences at the 5% level except for the Spanish sample, where the critical ratios for the four MLA skills tests were, respectively, -2.68**, -9.43**, -2.37*, and -2.81** ($**p < .01$, $*p < .05$, the negative sign indicating that the New York sample mean was smaller). The poorer showing of the New York sample in Spanish may reflect the kind of regional difference mentioned in Chapter V, Section 3 where it was found that mean scores of regions were somewhat correlated with the popularity of the language. Spanish is not as popular in New York state as it is in some other regions, and it is not as popular as French or Russian. Furthermore, the national sample may have included more native or near-native speakers of Spanish than the New York sample. (Both Tables 4.1 and A.4 are based on both "regular" and "odd" cases, i.e., native speakers and other special cases are not excluded.) There was only one instance of a highly significant difference in standard deviations, for the French Speaking score, where the New York sample was much more variable ($F_{329, 1169} = 1.71$, $p < .01$). This could possibly have been due to greater variation in rating standards at the time the New York group tapes were evaluated. The comparative values for the skewnesses of the score distributions are not inconsistent between the two samples.

Because the New York pilot study samples were in general quite similar to the national study samples, it was not thought worthwhile to do numerous detailed analyses of the New York samples. In any event, the small numbers of cases available in the New York samples would often have precluded such analyses. The remainder of this Appendix A is devoted to the reporting of a number of comparative analyses of possible interest.

Factors Associated with the Attainment of Foreign Language Skills

Data collected in the New York pilot study did not conveniently permit the breakdown of the sample by vocational plans in the same way as the national sample. However, Table A.5, analogous to Table 6.12, suggests that those planning to teach at the elementary school level were generally inferior, on the average, to those planning to teach at other levels, and that students planning to teach at the college level were generally superior. Only the latter finding agrees with one in the national study; in the national study, the elementary school teaching aspirants were not significantly different from the secondary school aspirants. Table A.5 agrees with data from the national study in showing that students planning to use foreign language skills in government or business work were representative of the total tested group.

In Table A.6, there are shown the numbers and percentages of "regular" cases who began the study of their major language in elementary school, high school, or college. These data are entirely consistent with those presented in Table 7.1, except that if anything, there are relatively fewer college starters in the New York samples. Table A.7 compares the mean scores on the skills tests of these various groups, as did Table 7.2 for the national study, and shows very much the same trends: that those who started in elementary school attained higher scores than those who started in high school, and that the latter, in turn, performed better than those who started in college. The trends were highly significant in French and Spanish groups. In the German groups, the trend was confirmed, although because of small numbers of cases it was not significant; it is noteworthy that in the New York samples the directionality of the trend for German was the same as for the French and Spanish groups, whereas it was reversed in the national sample.

Table A.4

Data from MLA Skills Score Distributions, New York Pilot Study
(All Cases Complete on 4 MLA Skills Tests)

Statistic	Listening	Speaking	Reading	Writing
FRENCH (N = 330)				
Max. Possible Score (Form A)	56	125	70	80
Highest Obtained Score	56	116	70	72
Median (Estimated)	45.0717	72.2625	49.4224	47.5869
Mean	43.9727	73.7758	49.3030	47.6212
Lowest Obtained Score	24	39	29	24
Chance Score	29	--	32.5	--
Standard Deviation	7.6298	12.4732	8.9696	10.0589
g_1 (Measure of Skewness)	-0.4321**	0.3640**	-0.0399	0.0102
GERMAN (N = 66)				
Max. Possible Score (Form A)	56	125	70	80
Highest Obtained Score	56	119	68	78
Median (Estimated)	45.9942	91.7345	54.8062	58.4458
Mean	45.0455	91.7879	53.2879	55.4091
Lowest Obtained Score	28	57	33	22
Chance Score	29	--	32.5	--
Standard Deviation	7.9304	16.3248	10.0676	14.7391
g_1 (Measure of Skewness)	-0.3589	0.0098	-0.4524	-0.6181*
RUSSIAN (N = 42)				
Max. Possible Score (Form A)	56	125	70	80
Highest Obtained Score	55	121	69	79
Median (Estimated)	42.2110	74.5163	41.4128	60.7136
Mean	42.5714	79.1667	43.3571	59.7857
Lowest Obtained Score	33	48	23	37
Chance Score	29	--	32.5	--
Standard Deviation	6.3401	16.3924	13.3467	11.5995
g_1 (Measure of Skewness)	0.1706	0.8511*	0.4370	-0.2400
SPANISH (N = 180)				
Max. Possible Score (Form A)	56	125	70	80
Highest Obtained Score	56	116	67	77
Median (Estimated)	43.6844	72.1865	45.9639	53.1280
Mean	43.4778	74.7556	46.9722	52.9333
Lowest Obtained Score	27	48	30	28
Chance Score	29	--	32.5	--
Standard Deviation	6.8138	13.6844	8.3762	11.0798
g_1 (Measure of Skewness)	-0.0910	0.5632**	0.3611*	-0.0527

**p < .01, *p < .05 for hypothesis of $g_1 = 0$ and normality of distribution.

Table A.5

Means and S.D.'s of MLA Skills Test Scores by Vocational Plans (Sexes Combined), and by Major Language, New York Pilot Study Sample

		FRENCH					GERMAN					RUSSIAN					SPANISH				
		N	L	S	R	W	N	L	S	R	W	N	L	S	R	W	N	L	S	R	W
Plans to teach																					
-at elementary school level	$\bar{X}$	31	38.4	68.8	41.2	40.4	2	Not indicated				1	Not indicated				10	37.4	68.7	40.3	45.2
	σ		6.9	8.2	5.6	7.5												4.2	9.1	4.2	6.1
			--	-	--	--												--		--	-
-at secondary school level	$\bar{X}$	190	43.5	72.9	48.7	46.8	26	42.7	85.3	50.4	52.8	13	40.2	77.2	39.7	56.3	106	41.9	73.0	45.6	49.5
	σ		7.1	11.8	8.5	9.1		7.3	14.0	9.6	13.8		5.9	12.7	10.2	11.5		6.6	12.2	8.3	10.3
																		-			--
-at college level	$\bar{X}$	52	46.9	77.3	53.5	51.7	21	44.1	92.0	53.1	55.1	9	Not indicated				32	45.1	79.1	50.1	57.2
	σ		7.4	12.0	9.0	9.4		8.0	14.7	9.3	13.9							6.7	13.7	7.6	11.1
			+	++	++	++													++	++	++
Plans to use FL																					
-in government work	$\bar{X}$	49	44.1	74.7	50.2	47.9	11	44.0	88.5	52.6	54.6	8	Not indicated				30	45.1	76.9	48.5	53.9
	σ		7.6	10.9	8.7	8.7		8.6	19.6	10.7	14.2							5.8	14.3	7.7	11.1
-in business	$\bar{X}$	55	44.0	74.9	50.3	48.3	12	45.3	91.6	53.4	53.3	2	Not indicated				45	42.7	74.5	45.2	48.9
	σ		8.0	12.0	8.6	10.2		8.9	16.9	11.0	14.9							6.6	13.7	7.8	11.0
																					-
TOTAL GROUP	$\bar{X}$	301	43.6	72.9	48.9	47.2	54	43.4	88.5	51.4	52.9	27	Not indicated				166	43.1	73.8	46.4	52.2
	σ		7.6	11.8	8.8	9.7		7.8	15.5	9.8	14.2							6.7	12.6	8.0	10.7

++ Significantly ($p < .01$) superior to remainder of total group
+ " ($p < .05$) " " " " " "
- " ($p < .05$) inferior " " " " " "
-- " ($p < .01$) " " " " " "

Table A.6

Numbers and Percentages of "Regular" Cases, by Time of Beginning FL Study, by Language (Excluding Italian) (All cases complete on four MLA Skills Tests) Data from Institutions in the State of New York

Time of Beginning	Language								Total	
	French		German		Russian		Spanish			
	N	%	N	%	N	%	N	%	N	%
"Grade school"	53	17.5	1	1.8	0	0.0	14	8.4	68	12.7
High school	231	76.2	31	54.4	2	20.0	131	78.9	395	73.7
College	19	6.3	25	43.9	8	80.0	21	12.7	73	13.6
Total	303		57		10		166		536	

Table A.7

Means and S.D.'s of MLA Skills Test Scores for Groups Starting FL Study at Different Educational Levels, New York Pilot Study Samples (All "Regular" Cases Complete on MLA Skills Test Scores)

Time of Beginning	N	Listening		Speaking		Reading		Writing	
		$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ	$\bar{X}$	σ
				FRENCH					
Grade School	53	46.11	7.60	78.42	10.94	51.40	7.70	50.02	8.94
High School	231	43.34	7.49	72.24	11.63	48.77	8.97	46.91	9.73
College	19	40.74	6.15	65.95	9.81	44.21	5.90	43.74	9.12
Total	303								
F		4.52*		10.01**		5.03**		3.62*	
				GERMAN					
Grade School High School	32	44.25	8.31	88.69	17.22	52.19	10.75	53.75	15.04
College	25	42.14	6.62	88.32	12.27	50.36	7.89	51.73	12.34
Total	57								
F		0.95		0.01		0.45		0.27	
				SPANISH					
Grade School	14	46.79	5.70	84.14	9.87	51.36	7.62	58.86	9.43
High School	131	43.14	6.56	72.91	12.32	46.24	7.89	52.08	10.56
College	21	40.38	6.61	72.43	12.11	44.29	7.24	48.62	9.77
Total	176								
F		4.02*		5.50**		3.57*		4.07*	

$^{**}p < .01$; $^{*}p < .05$

(It will be recalled, however, that for the national sample the discrepancy seemed to be explained by the fact that for some reason the German elementary and secondary school starters had had less total exposure to the language than those who started in college.)

Likewise, in the New York samples the amount of time spent abroad manifested the same strong association with test scores, as shown in Table A.8, as it did in the national sample (Table 7.9). (No cross-classification with time of beginning foreign language study was made for the New York study samples because of small numbers of cases.)

Table A.8

Mean Listening Test Scores (and N's) by Time Spent Abroad, New York Pilot Study Samples

	French	German	Spanish
Year Abroad	50.57 (49)	48.67 (9)	49.80 (20)
Summer school abroad	46.77 (48)	(3)	44.92 (25)
Toured abroad	44.40 (45)	46.10 (10)	44.48 (25)
Never abroad	40.29 (159)	40.56 (32)	40.86 (96)
TOTAL	43.61 (301)	43.39 (54)	43.10 (166)

Finally, regression analyses of New York pilot study data exhibited patterns of behavior in the predictors that were highly similar to those found for the national sample. Table A.9 gives beta-weights and multiple correlations for a set of predictors that includes the MLAT total score and several measures of the amount of formal training. It is roughly comparable to the regression analyses shown in the second set of data presented in Tables 8.2, 8.3, 8.4, and 8.5, except that it does not include Time Abroad. Save for this omission, the multiple correlations obtained for the New York study would probably have been as great in magnitude as those for the national study.

Table A.9

Multiple Regression Analyses for the MLA Skills Tests (Beta Weights and Multiple R's), New York Pilot Study Samples

Predictors \ Language	Listening Test				Speaking Test				Reading Test				Writing Test			
	F	G	R	S	F	G	R	S	F	G	R	S	F	G	R	S
MLAT-Total	.32**	.20	-.35	.21**	.09	.05	-.17	.10	.39**	.29*	-.25	.29**	.40**	.28*	-.18	.30**
Sem. G. S.	.12*	.38**	.30	.18*	.21**	.28	.04	.28**	.08	.38*	.26	.23**	.11	.31*	.44*	.22**
Sem. H. S.	.14*	.42**	.29	.05	.10*	.11	.42	.00	.12*	.29	.40	.04	.12*	.21	.41*	.05
Hrs. Coll Lang.	.13*	.33*	.09	-.12	.03	.21	.10	-.09	.06	.19	.05	-.05	.11	.16	.45*	-.06
Hrs. Coll. Lit.	.32**	.33*	.04	.32**	.26**	.25	.09	.11	.31**	.24	.04	.36**	.31**	.30*	.18	.32**
Hrs. Misc. Courses	.06	-.08	.24	.30**	.07	-.02	-.06	.27**	.08	-.07	.18	.30**	.03	-.11	.12	.27**
Multiple R	.50**	.60**	.53	.50**	.39**	.38	.46	.42**	.53**	.53*	.51	.54**	.53**	.52*	.66*	.52**
N	241	50	29	143	241	50	29	143	241	50	29	143	241	50	29	143

**p < .01

*p < .05

Appendix B

PJ-251
Form 10 (revised)

Bureau of the Budget No. 51-R454.1
Approval expires December 31, 1965

STUDY OF FOREIGN LANGUAGE MAJORS
IN UNITED STATES COLLEGES AND UNIVERSITIES, 1965
Conducted by Harvard University under NDEA Title VI
contract with the U.S. Office of Education

QUESTIONNAIRE FOR FOREIGN LANGUAGE MAJORS

TO THE STUDENT: This questionnaire has been devised to obtain essential background information to be analyzed in conjunction with the nationwide testing project in which you are participating.

This testing project is being supported by the United States Office of Education in its effort to help improve the quality of foreign language instruction in our nation's schools, colleges, and universities. Your careful attention to this questionnaire will be a service not only to yourself but also to future students of foreign language in our country.

Please follow the step-by-step directions closely in order to insure completeness and accuracy. The use of a pencil is recommended so that you can correct mistakes more easily.

Your answers to this questionnaire will be treated in confidence. However, your test results will be sent you in September or October, 1965, at the address you specify below.

Institution Code ☐ ← Please print the two code numbers given you by the test administrator in the appropriate boxes: → Student Code ☐

PART ONE

Directions: Please fill in the blanks by printing clearly or checking the proper answer for each of the items below:

1) Name ____
(last) (first) (initial)

Permanent address to which your test results may be mailed:

____
(no.) (street) (city) (state)

2) Sex: ☐$_{1}$ M ☐$_{2}$ F

Age: _____

3) I expect to graduate in (month) ______________ (year) __________ from (name of institution) ______________________________.

4) In the boxes below, check the one language in which you are being tested:

01 ☐ FRENCH
02 ☐ SPANISH
04 ☐ ITALIAN
08 ☐ GERMAN
16 ☐ RUSSIAN

This is the TEST LANGUAGE referred to elsewhere in this questionnaire. It is assumed that this language is a language in which you are "majoring" or "concentrating." If you are officially "majoring" in two or more languages (do not count minors), print their names here: ______________________________.

5) Write "T" in the box representing your TEST language. In each of the other boxes, write the approximate number of years of study you have had in the given language, in grade school, high school and college combined. Report to nearest 1/2 year, e.g. 1/2, 1, 2 1/2. Write "O" if you have taken no courses. Make sure that you have marked each box.

01 ☐ FRENCH
08 ☐ GERMAN
02 ☐ SPANISH
16 ☐ RUSSIAN
04 ☐ ITALIAN
☐ OTHERS (specify) ______________.

6) Have you ever taken previously any of the MLA Proficiency Tests for Teachers and Advanced Students? (These are the tests being given under the present project)

☐$_{1}$ NO ☐$_{2}$ YES (If "Yes," complete items A and B below):

A) Which language test(s) did you take? Check the language(s) involved:

☐$_{01}$ FRENCH ☐$_{02}$ SPANISH ☐$_{04}$ ITALIAN ☐$_{08}$ GERMAN ☐$_{16}$ RUSSIAN

B) Which particular test(s) did you take? If these were in more than one language, explain below:

☐$_{1}$ LISTENING ☐$_{2}$ SPEAKING ☐$_{3}$ READING ☐$_{4}$ WRITING ☐$_{5}$ CIVILIZATION AND CULTURE
☐$_{6}$ APPLIED LINGUISTICS ☐$_{7}$ PROFESSIONAL PREPARATION

____

____

PLEASE TURN THE PAGE

Appendix B (continued)

PAGE 2

This page is for reference as you complete the table on page 3.
Be sure to read the Directions on that page before you start work on the chart.

A **TYPE OF COURSE**

1 - any course at grade school level
2 - any high school course
3 - beginning or introductory college language course. A language (as opposed to literature) course is a course which may involve some reading, but is primarily concerned with basic grammar study, listening and speaking.
4 - "intermediate" language course at the college level
5 - "advanced" language course at the college level. Include here conversation courses, phonetics courses, advanced grammar or writing courses.
6 - literature courses at the college level. A literature course is devoted primarily to the reading of texts in the foreign language (not in translation), rather than the study of the language as such.
7 - any courses not covered by the above

B **DURATION (semesters or quarters)**
C

Use column B or C depending on whether the school was a semester or quarter system. If a high school or grade school course, consider a half year as 1 semester and a full year as 2 semesters. Consider a college summer course as 1 semester.

D **CLASS HOURS PER WEEK**

Write in the number of classroom hours per week for this course. A single class period of 40 minutes to 1 hour should be considered an "hour" in adding up the total per week.

E **SIZE OF CLASS**

Mark the estimated number of students there were in this class (under one teacher) according to the following scale:

1 - (fewer than 10) 2 - (10 to 19) 3 - (20 to 29) 4 - (30 or over)

F **TEACHER'S CLASSROOM LANGUAGE**

1 - the teacher spoke almost entirely in English except when reading foreign sentences or covering assigned drills
2 - the teacher occasionally would discuss things using the foreign language, converse with students in the foreign language
3 - during class periods, the teacher spoke the foreign language almost exclusively, using English as little as possible

G **STUDENTS' CLASSROOM LANGUAGE**

1 - speaking in English was the general rule, except for some short periods of conversation in the foreign language
2 - during class periods, the students were required to speak in the foreign language; only occasionally would English be allowed

H **LANGUAGE LABORATORY USE**

1 - the use of a language laboratory was not involved in this course
2 - a language laboratory was used on an informal or occasional basis
3 - the use of a language laboratory was an important and integral part of this course

I **TEACHER'S PRONUNCIATION**

1 - the classroom teacher was not a native speaker of the foreign language and pronounced it with a definitely non-native accent
2 - the classroom teacher had a somewhat non-native accent
3 - the classroom teacher was a native speaker of the foreign language, or had a "native" accent

J **NATIVE INFORMANTS** (native speakers of the language who assist the teacher by conducting drills or conversation sessions)

1 - the students did not meet with native informants in connection with this course
2 - the students met with native informants on a regular basis for language practice

K **FINAL EXAMINATION (Listening):** If the final examinations in this course involved (or will probably involve, in the case of current courses) the students' listening to statements or questions spoken in the foreign language, mark "1." If not, mark "0."

L **FINAL EXAMINATION (Speaking):** If the final examinations in this course involved or will involve speaking to the teacher (or a recorder) in the foreign language, mark "1," If not, mark "0."

M **FINAL GRADE:** My grade in this course was (or probably will be): (If course lasts more than 1 semester, give average grade as closely as possible:)

0 - Fail 1 - about "D" (or equivalent) 2 - about "C" (or equivalent) 3 - about "B" (or equivalent) 4 - about "A" (or equivalent)

Appendix B (continued)

PART TWO - COURSES IN MAJOR LANGUAGE

Directions: In the table below on "Courses in Major Language," the 25 rows are lines on which we would like you to answer a number of questions about each of the formal courses which you have taken in your MAJOR language (the language you will be tested on) from grade school up to and including the present time.

A formal course is any course offered in a regular school system and for which school credit is given. Formal courses may include summer, night or correspondence courses, provided they were taken for school credit. Any courses which you took "on your own," including courses which you audited, are not considered formal courses and should not be included here.

Space has been provided in the left-hand columns of the table to list each course, together with its grade level (for example, 7th grade, 1st semester fresh.), and the INSTRUCTOR'S NAME (if known). The course title can be approximate (e.g., Begin. French; 18th Cent. Lit., etc.).

What you write in these first three columns is for your own use to help you remember the courses, so your indications may be brief. Please be sure, however, that you list every course and that the courses are in chronological order, beginning with the earliest courses and ending with the ones you are currently taking. (In case two or more courses were taken simultaneously, list in any convenient order.)

After listing the courses, you will be ready to fill in the requested information for the rest of the table. The answer in each case will be some number, either a number requiring no code (for example, number of classroom hours per week), or a number corresponding to one of the coded answers given in the "Answer Key" provided on the facing page.

The completion of this table is the major portion of the questionnaire. Spend sufficient time on it and complete it as accurately as possible. If you find it absolutely impossible to mark a particular item, write in "DK" (don't know), but as far as possible try to answer all the questions.

COURSES IN MAJOR LANGUAGE

GRADE LEVEL	APPROX. COURSE TITLE	INSTRUCTOR		A) TYPE OF COURSE	B) DURATION (Semesters) (mark only one in each case)	C) DURATION (Quarters)	D) CLASS HOURS PER WEEK	E) SIZE OF CLASS	F) TEACHER'S CLASS LANGUAGE	G) STUDENTS' CLASS LANGUAGE	H) LANGUAGE LABORATORY USE	I) TEACHER'S PRONUNCIATION	J) NATIVE INFORMANTS	K) FINAL EXAM (Listening)	L) FINAL EXAM (Speaking)	M) FINAL GRADE
			1													
			2													
			3													
			4													
			5													
			6													
			7													
			8													
			9													
			10													
			11													
			12													
			13													
			14													
			15													
			16													
			17													
			18													
			19													
			20													
			21													
			22													
			23													
			24													
			25													

TURN TO PART III ON PAGE 4.

Appendix B (continued)

PART THREE PAGE 4

Directions: Check the following items as appropriate.

1) For what reason(s) did you decide to concentrate in the language in which you are being tested? Check as many as apply:

- ☐ I expect to teach a foreign language or languages. If so, check the level or levels at which you plan to teach:
  - ☐ at the elementary school level
  - ☐ at the secondary school level
  - ☐ a college, university, or graduate school
  - ☐ other (specify)__________________
- ☐ I expect to use my foreign language knowledge in my future employment or profession. If so, specify the kinds of work you are contemplating:
  - ☐ Linguistics
  - ☐ Business activity (importing, foreign trade, etc.)
  - ☐ Government work (foreign service, United Nations, A.I.D., etc.)
  - ☐ Military service
  - ☐ Work as a translator or interpreter
  - ☐ Other professional work (medicine, science, library work, etc.)
- ☐ I expect to use my foreign language knowledge in travel or study abroad.
- ☐ I majored in a foreign language mainly for interest and enjoyment.
- ☐ I majored in a foreign language because I am particularly interested in the literary and cultural aspects of the language.
- ☐ I am interested in intercultural understanding and communication.
- ☐ No particular reason other than the fact that I needed a major subject.
- ☐ Other reasons (specify) ______________________________.

2) In view of the reason(s) for selecting your language major indicated above, please estimate the extent to which skill in each of the four major language areas (listening, speaking, reading, writing) would be of importance to you. Please mark a "check" for each of the four skills.

	little importance	some importance	great importance
Listening:	1☐	2☐	3☐
Speaking:	1☐	2☐	3☐
Reading:	1☐	2☐	3☐
Writing:	1☐	2☐	3☐

3) Are you planning to qualify for a license or certificate for teaching in public schools?

☐$_1$ No ☐$_2$ Yes

4) Considering only your major language, have you taken informally (not for credit) any course, or audited any course, in that language (whether summer, night, special or other)?

☐$_1$ No ☐$_2$ Yes

5) What courses have you taken in Latin or Greek?

☐$_1$ None ☐$_1$ One or more in high school ☐$_1$ One or more in college

6) For purposes of this question, assume that "summer study abroad" in the major language would mean that you travelled to a country where your major language is spoken natively and took a regularly scheduled course or courses in your major language during that time. (These courses are not necessarily school credit courses).

Have you had such summer study abroad?

☐$_1$ No ☐$_2$ Yes

7) For this question, assume that "school year study abroad" would mean that you spent a regular school year in a country where your major language is spoken natively and took a series of courses in your major language during that time.

Have you had such a school year study abroad?

☐$_1$ No ☐$_2$ Yes

8) Skip to question 9 if you answered yes to either question 6 or 7. If your answer to both questions 6 and 7 was no, mark the appropriate answer to the question below:

Have you travelled as a tourist in the major-language country and/or had a program of self-study in the major-language country?

☐$_1$ No ☐$_2$ Yes

9) Do you and/or your parents speak your major language at home?

☐$_1$ No ☐$_2$ Yes, occasionally ☐$_3$ Yes, frequently

10) Have you had occasion to use your major language in a job situation, with foreign friends, or in some other extracurricular situation?

☐$_1$ No ☐$_2$ Yes, to a moderate extent ☐$_3$ Yes, extensively

11) Independent of and in addition to reading required in courses, have you had any other reading experiences in your major language? Check the one sentence which best describes your experiences.

☐$_1$ I have not done any reading in my major language except for material required by or involved in course work.

☐$_2$ I have read, independently of course work, a few pages of material in my major language.

☐$_3$ I have read, independently of course work, one or two books (or the equivalent) in my major language.

☐$_4$ I have read, independently of course work, three or more books (or the equivalent) in my major language.

END OF QUESTIONNAIRE

Appendix C

OE-4224-1 (5-64)
PJ-251
Form 11

Bureau of the Budget No. 51-R454
Approved: April 8, 1964
Approval expires: March 1, 1965

STUDY OF FOREIGN LANGUAGE MAJORS
IN UNITED STATES COLLEGES AND UNIVERSITIES, 1964

conducted by Harvard University under NDEA Title VI contract with the U.S. Office of Education

QUESTIONNAIRE FOR CHAIRMEN OF FOREIGN LANGUAGE DEPARTMENTS

TO THE CHAIRMAN: In order to obtain a certain amount of general information relevant to the organization and conduct of the foreign language programs in your department, we would greatly appreciate your answers to the questionnaire below. In some cases, the appropriate answer will consist simply of a number or check mark. In others, it will be necessary to describe at some length certain details of your language program. Although a certain amount of writing space has been provided in each instance, you may find it necessary to append a separate sheet for answers to some questions. If this is necessary, please number these questions as they are numbered in the questionnaire.

We would like to take this opportunity to thank you in advance for your attention to this questionnaire and to express our appreciation for your participation in the project as a whole.

NAME OF INSTITUTION:____________________

If this department is at a particular branch of the institution, please indicate:__________

OFFICIAL NAME OF DEPARTMENT:____________________

NAME OF DEPARTMENT CHAIRMAN:____________________

INSTITUTION CODE
(do not fill in)

DEPARTMENT CODE
(do not fill in)

Please check all the languages for which courses are given in your department. (Check all that apply)

1[ ] FRENCH	2[ ] SPANISH	4[ ] ITALIAN	8[ ] GERMAN
16[ ] RUSSIAN	32[ ] OTHERS (Specify)__________		

Please turn to page three.

Appendix C (continued)

1) At your institution, when do students normally elect their "major" or subject matter field of concentrated study? (Circle one number)

	FIRST YEAR	SECOND YEAR	THIRD YEAR	OTHER (describe)
If semester system:	1 2	1 2	1 2	________
If quarter system:	1 2 3	1 2 3	1 2 3	________

2) What guidance, if any, is available to students in helping them make decisions as to "major" or field of concentrated study? (Check all that are applicable)

1[ ] None; student makes decision without help from faculty or staff.
2[ ] Help is available from a faculty advisor assigned to the student.
3[ ] The student may seek the advice of members of the foreign language department(s).
4[ ] Assistance is available from a counselling and testing bureau on the campus.
5[ ] Other (specify)________

3) What requirements, if any, must undergraduates meet before being accepted as "majors" in your department? (If applicable, describe requirements in terms of academic record, scores on aptitude and proficiency tests, and the like.)

4) Please describe the specific foreign language DEGREE requirements for an A.B. degree with a major in your department. Include in your answer how many courses of each of several types (language, literature, composition, conversation, culture and civilization, etc.) are required. Please indicate also any use of proficiency tests to check foreign language achievement.

5) Does your department offer programs leading to any degree beyond the A.B.? (Check all that are applicable)

0[ ] None.
1[ ] M.A.
2[ ] Ed.M.
3[ ] Ph.D.
4[ ] Ed.D.
5-9[ ] Other (Specify)________

Please turn to next page.

Appendix C (continued)

4

6) Please indicate the number of "majors" currently enrolled in your department, by class and by language. For languages not covered in your department, write N.A. ("not applicable").

	FRENCH	SPANISH	ITALIAN	GERMAN	RUSSIAN	INTERDEPARTMENTAL (describe)
class of 1964 (seniors)						
class of 1965 (juniors)						

7) Please indicate with a check [✓] which of the following periodicals are subscribed to by either the main institution library or by your departmental library: (check one or more)

1[ ] Georgetown Monograph Series in Language and Linguistics
2[ ] Language (Linguistic Society of America)
3[ ] Language Learning
4[ ] Modern Language Journal
5[ ] PMLA
6[ ] None of these

Now consider each group of periodicals pertaining to the language(s) covered in your department, and make checks to indicate which are subscribed to by either the main institution library or by your departmental library:

FRENCH

11[ ] Le Français Moderne
12[ ] French Review
13[ ] Romance Philology
14[ ] Romania
10[ ] None of these

SPANISH

21[ ] Hispania
22[ ] Revista de Filología Española
23[ ] Revista Hispanica Moderna
24[ ] Thesaurus - Boletín del Instituto Caro y Cuervo
20[ ] None of these

RUSSIAN

161[ ] Revue des Études Slaves
162[ ] The Slavic and East European Journal (AATSEEL)
163[ ] Slavonic and East European Review (London)
164[ ] Zeitschrift für Slavische Philologie
160[ ] None of these

ITALIAN

41[ ] Italica
42[ ] Lingua Nostra
43[ ] Rendiconti della Reale Accademia d'Italia
44[ ] Studi di Filologia Italiana
40[ ] None of these

GERMAN

81[ ] German Life and Letters (London)
82[ ] German Quarterly
83[ ] Germanic Review
84[ ] Monatshefte
80[ ] None of these

8) In the rows of the following table are categorized various types of courses which often figure in undergraduate foreign language instruction. For each language which your department offers, please indicate the number of semester-length courses (see footnote if your institution is on the quarter system) which are offered to undergraduates in each of these categories. If no courses are offered to undergraduates in a particular category, mark "0". Please be careful to define the number of courses involved on a semester-length basis. A course which continues for a full year would count as two semester-length courses for purposes of this table. Please be careful to exclude courses which are not available to undergraduates.

	FR	SP	IT	GER	RUS
BEGINNING OR INTRODUCTORY LANGUAGE COURSES: Language courses (as opposed to "literature" courses) may involve one or more of the four basic skills (listening, speaking, reading, writing) and introduce the student to the basic structure and vocabulary of the language.					
INTERMEDIATE LANGUAGE COURSES: Intermediate language courses presuppose some prior knowledge of the language and continue to stress the further development of one or more of the basic skills. Textual materials are introduced chiefly to give the student practice in handling the language.					

A note for institutions operating on the quarter system: Institutions which use the quarter system should fill out the table on the basis of the number of quarter-length courses offered, and this box should be checked: [ ]QUARTER SYSTEM

Appendix C (continued)

	FR	SP	IT	GER	RUS
COMPOSITION COURSES: Courses usually available only to more advanced students and which concentrate on the rather intensive development of writing skill. (see footnote)					
CONVERSATION COURSES: Courses usually available only to more advanced students and which concentrate on the development of a rather high degree of speaking fluency. (see footnote)					
PHONETICS COURSES: "Phonetics" courses are distinguished from "conversation" courses in that the primary instructional emphasis is on the acquisition of a native-like pronunciation and intonation, with somewhat lesser attention paid to the development of verbal fluency as such.					
ADVANCED GRAMMAR OR SYNTAX COURSES:					
LITERATURE COURSES: "Literature" courses are devoted primarily to the reading of texts in the foreign language (not in translation), rather than the study of the language as such. Literature courses include the so-called "survey" courses, as well as the study of particular epochs, genres, schools, or authors.					
SPECIAL READING COURSES: Include here any reading courses which are offered for such specifically practical purposes as the acquisition of a reading knowledge in scientific or technical fields. Include here also any reading courses (usually of a terminal nature) which, although they may involve more literary readings, aim primarily at the development of basic reading skills rather than the appreciation of the particular works involved.					
CULTURE AND CIVILIZATION COURSES:					
COURSES IN METHODS OF TEACHING FOREIGN LANGUAGES:					
DIACHRONIC OR HISTORICAL LINGUISTICS COURSES:					
SYNCHRONIC OR DESCRIPTIVE LINGUISTICS COURSES:					
The remaining spaces should be used to describe and number any courses offered to undergraduates by your department which are not adequately covered by the above categories:					
______________________					
______________________					
______________________					
______________________					
______________________					

COMBINATION COURSES in composition and conversation: If your institution offers a combined composition-conversation course, please place numbers or fractional numbers in both the composition and conversation course blanks to reflect the relative emphasis devoted to these respective activities.

Please turn to next page.

Appendix C (continued)

9) Please check in the table below which of the following activities are part of your foreign language program and the extent to which they are required of foreign language majors at some time during the undergraduate course of studies.

YEAR ABROAD: a program providing for an academic year abroad (e.g., junior year abroad)
SUMMER ABROAD: a program whereby a student goes to summer school abroad and/or pursues a summer study tour abroad.
LANGUAGE HOUSE: a separate residential unit or section of a dormitory in which a particular foreign language is the only language of communication permitted.
LANGUAGE TABLES: separate tables or cafeteria sections where a particular foreign language is the only language of communication permitted.
LANGUAGE CLUBS: extracurricular organizations which promote interest in a particular foreign language and in the culture and history of the countries where the language is spoken.
PLEDGE: a personal commitment to speak only in the foreign language in designated places and at designated times.

__KEY TO THE TABLE__ - N.A. (not available) V (voluntary) R (required)

Please place check marks in __all__ applicable categories and for __all__ applicable languages:

	__FRENCH__			__SPANISH__			__ITALIAN__			__GERMAN__			__RUSSIAN__		
	NA	V	R	NA	V	R	NA	V	R	NA	V	R	NA	V	R
YEAR ABROAD:															
SUMMER ABROAD:															
LANGUAGE HOUSE:(during (during school year)															
LANGUAGE HOUSE: (during the summer)															
LANGUAGE TABLES:															
LANGUAGE CLUBS:															
PLEDGE:															

10) Which of the following statements best describes the department policy in designing and teaching the "basic skills" courses (normally, the two-year sequence which attempts to bring students from virtually no knowledge of the language up to a point where they are sufficiently proficient to handle the language in advanced courses in literature, culture, and civilization, linguistics, composition, and so forth.)

Please check __one__ and fill in the blanks appropriately:

A. GENERAL. __In basic skills courses__:

1[ ] Students are divided into sections and each section is taught by a single instructor.

a) Average number of students per section:________

b) Number of hours per week each section meets:________

2[ ] Some instruction is given to students meeting in large groups and some in small sections taught by regular instructors (__not native informants__).

a) Average number of students in large sections:________

Number of hours per week large sections meet:________

b) Average number of students in small sections:________

Number of hours per week small sections meet:________

3[ ] Some instruction is given to students meeting in large groups and some in small sections led by a native informant for purposes of drill and conversation.

a) Average number of students in large sections:________

Number of hours per week large sections meet:________

b) Average number of students in small sections:________

Number of hours per week small sections meet:________

4[ ] Other (describe in terms comparable to the above):____________________

____

____

Appendix C (continued)

B. USE OF LANGUAGE LABORATORY: In conjunction with basic skills courses: (check one and fill in any blanks appropriately)

1[ ] No language laboratory facilities are available.

2[ ] Language laboratory facilities are available:

______hours per week for required work.

______hours per week at the student's option.

C. COURSE EMPHASIS: In basic skills courses: (check one)

1[ ] More emphasis is given to listening and speaking than to reading and writing.

2[ ] More emphasis is given to reading and writing than to listening and speaking.

3[ ] About the same emphasis is given to listening and speaking as to reading and writing.

4[ ] Our policy may be better described as follows:____

____

D. RELATIVE AMOUNT OF USE OF ENGLISH: (Using the following symbols, place one in each of the spaces in the appropriate columns of the table)

F - if the teachers customarily use English
O - if the teachers occasionally use English
R - if the teachers rarely use English

In basic skills courses, English is used (place one symbol in each space)

	FIRST YEAR	SECOND YEAR
1) to explain points of grammar	[ ]	[ ]
2) to explain the meaning of new words and phrases	[ ]	[ ]
3) to give general instructions to the student	[ ]	[ ]
4) other (specify)______________________________	[ ]	[ ]

E. ADDITIONAL COMMENTS: Space is provided for a description of any other aspects of departmental policy or practice with regard to the basic language courses which is not covered by the above.

____

____

____

____

____

____

____

____

____

____

PLEASE MAIL COMPLETED QUESTIONNAIRE (using accompanying reply envelope) to: College Foreign Language Testing Project, Longfellow Hall, Harvard University, Cambridge, Mass. 02138

Appendix D-1

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION
WASHINGTON, D.C. 20202

MEMORANDUM TO PRESIDENTS OF INSTITUTIONS OF HIGHER LEARNING

FROM: Kenneth W. Mildenberger KM
Director, Division of College and University Assistance

As you know, the National Defense Education Act of 1958 has stimulated considerable research and improvement in the teaching of modern foreign languages.

As a part of its program of support of research studies in language teaching, the U.S. Office of Education has contracted with Harvard University to administer a new series of foreign language proficiency tests to a representative group of foreign language majors in their senior year at colleges and universities throughout the Nation. I believe you will agree that a careful study of foreign language programs and students' achievements in them will be of great value to all administrators and departments concerned.

The plan calls for testing students in all institutions in the 50 percent sample to be studied. The testing program will be offered free of charge to all seniors in the institutions who are majoring in any one of the five languages (French, Spanish, Italian, German, Russian) for which tests are available. The tests are to be given early in 1965. The data will be used for statistical purposes only without identifying the participating individuals or institutions in the report of the survey findings.

Would you or your academic dean complete the enclosed Institution Information Form and send it as soon as possible to Harvard University at the address indicated on the form. This form provides space on which you may indicate whether your institution will participate; it also asks for certain information which is needed for the study regardless of whether your institution participates. If your institution will participate, will you appoint a coordinator to receive more detailed correspondence regarding the study and to supervise the actual administration of the tests. The coordinator could well be a chairman of a modern foreign language department, a person in charge of language laboratories, or a person in charge of a college testing service.

We feel this is an important project, and we very much hope your institution will be able to take part in it.

Enclosure

Appendix D-2

PJ-251
Form 2 (revised)
Nov. 5, 1964

COLLEGE FOREIGN LANGUAGE TESTING PROJECT
A Study of the Foreign Language Proficiency of Language Majors
Near Graduation from College
Conducted by Harvard University under contract with the U.S. Office of Education

MEMORANDUM DESCRIBING THE PROJECT

Under a research contract with the U.S. Office of Education, Harvard University is conducting a study in which a new series of foreign language achievement tests will be administered without charge to a large sample of foreign language majors graduating from United States colleges and universities in the spring of 1965.

THE TESTS

The tests to be used in this nationwide project are the Foreign Language Proficiency Tests for Teachers and Advanced Students in French, German, Italian, Russian, and Spanish. They were produced by the Modern Language Association of America under an NDEA Title VI grant, and are designed to measure proficiency in four skill areas: Reading, Writing, Listening, and Speaking, and three areas of knowledge relevant to teaching: Civilization and Culture, Applied Linguistics, and Professional Preparation. They represent the most comprehensive measures currently available at the intermediate and advanced levels of foreign language study.

ELIGIBILITY TO TAKE THE TESTS

The four skills tests will be administered to all senior-class students majoring in one of the five languages in which the tests are available. (Junior class students are not eligible.) The three additional subtests related to teaching will be given to students certified as completing teacher preparation programs. All students will furthermore be administered a short form of a modern language aptitude test as well as a brief questionnaire on the history of their study of foreign languages, their interests in foreign language study, and other relevant data.

PURPOSE OF THE TESTING PROGRAM

The testing program will not only serve a function of national importance but will also prove of benefit to both the participating institutions and the students involved. On the national level, results of this study will allow a detailed appraisal of the functional foreign language proficiency of the pool from which future teachers of foreign languages---for all educational levels---will come. For the institutions involved, such information will permit an evaluation of their own foreign language programs; and for the students themselves, participation in the study will provide an opportunity to assess their own abilities by reference to national norms.

THE TESTS ARE FREE OF CHARGE

There will be no charge to the participating institutions or to the students for the administration of the tests. (The full battery of MLA Foreign Language Proficiency Tests normally costs $13.50 per student; the partial battery, consisting of the four skill tests, costs $10.00 per student.) Score reports, together with test norms and other interpretative material, will be sent both to the individual students and to the appropriate language department chairmen by the early fall of 1965. All information collected will be treated as confidential and statistical data will not be identified with individual students, departments, or institutions in the published reports.

DETAILS OF TEST ADMINISTRATION

With respect to the details of the test administration, the president or academic dean of each participating institution has been asked to appoint a coordinator to whom the test booklets and other materials will be shipped in time for testing in March or April, 1965, and who will be directly responsible for the administration of the tests and the return of materials. Coordinators and any assisting proctors will receive nominal fees for their services. It is expected that the coordinator will be able to work closely with each department chairman in securing the greatest possible participation on the part of the students. The testing program is expected to have considerable appeal to students in view of the fact that achievement tests of the type to be administered are being increasingly used by graduate schools, business concerns, and other organizations as part of their application and placement procedures. Score reports for the different skill areas are likely to be useful to students who are contemplating graduate study in foreign languages or who plan to use foreign languages in some aspect of their future career.

It is estimated that the total testing time (including rest periods and time for instruction) will be close to four hours, except that an additional two hours will be required in the case of students taking the parts related to teaching.

Questions or comments in connection with the project may be referred to the address below:

College Foreign Language Testing Project
Longfellow Hall
Harvard University
Cambridge, Mass. 02138
(Telephone: Area Code 617, Un 8-7600, Ext. 3458)

Appendix D-3

OE - 4224C
PJ-251
Form 20

Bureau of the Budget No:51-R454.1
Approved: November 23, 1964
Approval expires:December 31, 1965

COLLEGE FOREIGN LANGUAGE TESTING PROJECT
A Study of the Foreign Language Proficiency of Language Majors near Graduation from College

INSTITUTION INFORMATION FORM

Directions: Please print or typewrite throughout. Return this form to: College Foreign Language Testing Project, Graduate School of Education, Harvard University, Cambridge, Massachusetts 02138, ON OR BEFORE DECEMBER 19, 1964

1) This item should be answered by all institutions:

NAME OF INSTITUTION ______________________________

WILL YOUR INSTITUTION PARTICIPATE THIS SPRING IN THE TESTING PROJECT? ☐ Yes ☐ No

Institutions not participating: Please answer the next item (item 2) and return this form using the reply envelope provided.
All participating institutions: Please answer all of the items below and on the reverse side, then return this form using the reply envelope provided.

2) LANGUAGES TAUGHT AT INSTITUTION; NUMBER OF ENROLLED SENIORS

The tests are given in five languages: French, Spanish, Italian, German, and Russian. Please use the boxes below to indicate for which of these languages you have an A.B. (or equivalent) degree program, and to report the exact number of seniors who are currently enrolled as "majors" or "concentrators" in each language and who will presumably be candidates for the degree at the end of the academic year 1964-65.

	FRENCH	SPANISH	ITALIAN	GERMAN	RUSSIAN
BACHELOR'S DEGREE PROGRAM:	☐ Yes [01] ☐ No	☐ Yes [02] ☐ No	☐ Yes [04] ☐ No	☐ Yes [08] ☐ No	☐ Yes [16] ☐ No
NUMBER OF SENIORS CURRENTLY "MAJORING" IN THE LANGUAGE:	_____	_____	_____	_____	_____

-- --

3) DESIGNATION OF COORDINATOR

The person named below has been appointed coordinator for the testing program and has agreed to serve in this capacity:

____
Title First name Initial Last name

4) MAILING ADDRESS OF COORDINATOR

All tests and other materials will be sent to the coordinator at this address. Please make sure that this address is sufficiently complete for mailing purposes.

____

____

____

TELEPHONE: Area Code_____ No.__________Extension_____ MORE ITEMS ON REVERSE SIDE

Appendix D-3 (continued)

5) NAMES AND ADDRESSES OF DEPARTMENT CHAIRMEN

The names and addresses of your foreign language department chairmen are needed for our records, not only in order for us to know what departments exist at your institution and which languages they cover, but also because it will later be necessary for us to collect certain information directly from these chairmen.

Please use as many as necessary of the boxes below to give the names and mailing addresses of the language chairman or chairmen responsible for each of the (test) languages taught at your institution. In the brackets at the left of each box, check the language or languages for which each chairman is responsible.

NAME OF CHAIRMAN

FRENCH [01] [ ]
SPANISH [02] [ ]
ITALIAN [04] [ ]
GERMAN [08] [ ]
RUSSIAN [16] [ ]

Title First name Initial Last name

MAILING ADDRESS

NAME OF CHAIRMAN

FRENCH [01] [ ]
SPANISH [02] [ ]
ITALIAN [04] [ ]
GERMAN [08] [ ]
RUSSIAN [16] [ ]

Title First name Initial Last name

MAILING ADDRESS

NAME OF CHAIRMAN

FRENCH [01] [ ]
SPANISH [02] [ ]
ITALIAN [04] [ ]
GERMAN [08] [ ]
RUSSIAN [16] [ ]

Title First name Initial Last name

MAILING ADDRESS

NAME OF CHAIRMAN

FRENCH [01] [ ]
SPANISH [02] [ ]
ITALIAN [04] [ ]
GERMAN [08] [ ]
RUSSIAN [16] [ ]

Title First name Initial Last name

MAILING ADDRESS

NAME OF CHAIRMAN

FRENCH [01] [ ]
SPANISH [02] [ ]
ITALIAN [04] [ ]
GERMAN [08] [ ]
RUSSIAN [16] [ ]

Title First name Initial Last name

MAILING ADDRESS

Deadline for return of form: DECEMBER 19, 1964 (An earlier return is requested if possible)

Appendix D-4

INSTRUCTIONS AND SUGGESTIONS
TO COORDINATOR FOR
ADMINISTERING TEST MATERIALS UNDER
THE COLLEGE FOREIGN LANGUAGE TESTING PROJECT
SPRING 1965

This leaflet is intended to provide a master information source for all activities and responsibilities of the institution coordinator in connection with the language testing project.

CHECK LIST OF MATERIALS

As a first step, would you please check the materials which you have received in this mailing against the list below.

IN MANILA ENVELOPE LABELED "MATERIALS FOR COORDINATOR"--

1) ONE COPY OF THIS LEAFLET
2) ONE COPY OF "ORDER FORM FOR ADMINISTRATION OF MLA FOREIGN LANGUAGE PROFICIENCY TESTS FOR TEACHERS AND ADVANCED STUDENTS UNDER THE COLLEGE FOREIGN LANGUAGE TESTING PROJECT"
3) ONE AIR MAIL BUSINESS REPLY ENVELOPE for immediate return of ORDER FORM to Harvard
4) ONE "APPLICATION FOR REIMBURSEMENT OF COORDINATORS AND ASSISTING PROCTORS"
5) ONE FIRST CLASS (NON-AIR MAIL) BUSINESS REPLY ENVELOPE for return of APPLICATION FOR REIMBURSEMENT form to Harvard following test administration.

IN MANILA ENVELOPE LABELED "MATERIALS FOR DEPARTMENT CHAIRMEN"--

1) from 1 to 5 copies of "MEMORANDUM DESCRIBING THE PROJECT"
2) an equal number of copies of the "QUESTIONNAIRE FOR CHAIRMEN OF FOREIGN LANGUAGE DEPARTMENTS"
3) an equal number of BUSINESS REPLY ENVELOPES for return of the QUESTIONNAIRE(s) FOR CHAIRMEN, to be returned by the department chairmen to Harvard as soon as conveniently possible.

FREE IN MAILING CARTON(s):

1) A sufficient supply of the booklet "A DESCRIPTION OF THE MLA FOREIGN LANGUAGE PROFICIENCY TESTS FOR TEACHERS AND ADVANCED STUDENTS" to distribute to each foreign language major graduating from your institution in the spring of 1965 in one of the five test languages.
2) An equal number of copies of the leaflet "A LETTER TO COLLEGE SENIORS GRADUATING IN THE SPRING OF 1965 WITH DEGREES IN FRENCH, ITALIAN, SPANISH, GERMAN, OR RUSSIAN." These leaflets are enclosed individually in the larger descriptive booklet.

IF ANY OF THE ABOVE MATERIALS ARE MISSING OR IN INSUFFICIENT SUPPLY, please place a collect call to College Foreign Language Testing Project, Area Code 607, UN 8-7600, Ext. 3459.

INFORMING STUDENTS ABOUT THE TESTING PROJECT

As soon as possible, arrangements should be made to distribute one explanatory booklet, and the accompanying enclosure, to every senior graduating in the spring of 1965 in one of the five test languages. The exact method to be used will vary depending on the size and organizational setup of the institution. In some cases, it may be possible for the coordinator to distribute these materials to the students personally. In other cases, chairmen of the language departments involved might undertake to make this distribution during class periods.

In any event, arrangements should be made so that all students eligible for testing receive these official explanatory materials as soon as possible.

COMPLETION OF TEST ORDER FORM

Among other things, the test order form requests information about

1) the test administration date chosen by your institution
2) the number of students to be tested in each language.

With respect to the TEST ADMINISTRATION DATE, each institution is free to choose the most convenient testing date, the chief restriction being that this date must fall within the period March 1 - May 8, 1965. An additional restriction is explained as follows: the testing date chosen for your institution must be at least one month later than the date on which the order form is sent to Harvard. Thus, for example, if an institution returns the order form on March 15, April 15 is then the earliest testing date which could be selected for that institution. Since testing dates in Late April or in May can be expected to conflict to some extent with vacation or examination periods at the institution, it is recommended that an earlier testing date be chosen if at all possible.

Regardless of the testing date actually chosen, it is urged that the coordinator make every effort to return the test order form to Harvard as soon as he can conveniently do so; this will enable the project staff to arrange for test shipments to all institutions more efficiently.

With respect to determining the NUMBER OF STUDENTS to be tested in each language, the following procedure is recommended:

All students have been urged, in the letter accompanying the explanatory booklet, to make a commitment to taking the test series on the testing date chosen by the institution. It would be well if the coordinator or the department chairmen could arrange to ask each student personally if he will participate in the project, and let the student know that his answer will affect the number of tests actually ordered, plans made for test administration, and so forth. It is not suggested that the students be formally required to take the tests, but any techniques which might reasonably be used to maximize student participation should be in order.

The MLA proficiency tests are given in two batteries: the Complete Battery (see order form) contains the four skill tests--Listening Comprehension, Speaking, Reading, Writing, and also the three teaching oriented tests--Applied Linguistics, Civilization and Culture, and Professional Preparation.

The PARTIAL Battery contains the four skill tests only.

The COMPLETE Battery of 7 tests may be administered only to those senior class students who are enrolled at your institution in a program which is expressly oriented toward

Appendix D-4 (continued)

the preparation of elementary or secondary school language teachers, and which usually involves the completion of a certain series of courses necessary for state certification purposes. It will not be possible to administer the complete battery of tests (that is, to give the three additional teacher preparation tests) to students who are following regular courses of study usually involved in the A.B. degree program in a particular foreign language. Further, no tests are to be administered to junior class students.

The coordinator is asked to determine and enter on the order form the exact number of students to be tested in each of the two categories (presumably, this will be the number of students who have agreed to participate in answer to the personal request suggested earlier).

It is important that only the number of tests actually used be ordered. Educational Testing Service normally includes a slight overage in test shipments so that such last-minute problems as a few additional test takers, a defective booklet, etc. can be handled adequately. It is thus not necessary to order additional tests for "safety's sake"; if student canvassing has been successfully conducted, the number of tests to be ordered should be specifically determinable.

The additional information required on the order form is largely self-explanatory. Please be sure, however, to check the type of recording tape materials which you will require for the Listening and Speaking tests; please also complete and sign the "Terms of Test Use" on the back of the order form.

Test order forms should be mailed to Harvard (not to ETS) using the air mail reply envelope provided.

COMMUNICATION WITH DEPARTMENT CHAIRMEN

The coordinator, in all probability, will already have been in contact with chairmen of the language departments for the five test languages (or for as many as are taught at the institution), in connection with various testing arrangements. We are asking at this time that two additional details be conveyed to the department chairmen:

First, they should be informed that rosters of test scores for all students in their departments will be sent to them in the early fall of 1965; along with the rosters will come norms and other materials allowing them to appraise their students' performance in the various language skills.

Second, we ask that you send one copy of the "QUESTIONNAIRE FOR CHAIRMEN" to each department chairman responsible for one (or more) of the languages in which the tests are actually to be given at your institution. According to our records, exactly the right number of questionnaires are included in this mailing; if additional questionnaires are needed, please inform us to this effect.

There is no formal deadline for the return of the QUESTIONNAIRE(s) FOR CHAIRMEN to Harvard, although we hope that they will be returned within approximately two weeks of receipt. Individual business reply envelopes are included with every questionnaire, so that

Appendix D-4 (continued)

each department chairman may mail the questionnaire directly to us following completion.

TEST ADMINISTRATION PROCEDURE

The exact details of test administration will vary from institution to institution, depending on both the total number of students to be tested and the time periods available for testing. However, certain general outlines may be suggested. Most students will require from 3 1/2 to 4 hours for the testing, divided as follows:

MLA Foreign Language Proficiency Tests (Listening Comprehension, Speaking, Reading, Writing) Approximate total time	2 1/2 hours
Student Questionnaire (approximate time)	1/2 hour-1 hour
Modern Language Aptitude Test (short form)	1/2 hour

This testing could occur in one session, either in the morning, the afternoon, or (if absolutely necessary) the evening (starting early).

Students taking the MLA Proficiency Tests relating to teacher preparation will take an additional session of tests lasting about 2 hours. This could be scheduled at any convenient time, either on the same day as the other session, or on a different day.

For the two tests requiring the use of tape recorder equipment (the Listening and Speaking tests), the capacity of the language laboratory (if one is used) must be taken into account in scheduling the tests. Regardless of the exact laboratory arrangements, any given student should be administered the Listening and Speaking tests in that order. (The order of administration for the pencil-and-paper tests--including the Modern Language Aptitude Test and the student questionnaire--may vary depending on conditions at the institution.)

The package of test materials shipped from ETS will include more specific administration details, including the official testing instructions for each test in the series.

RETURN OF TESTING MATERIALS

With the exception of the student questionnaires, which are to be returned to Harvard, all testing materials should be returned to ETS as soon as possible following administration, using the prepaid shipping system provided. Student questionnaires are returned direct to Harvard in special prepaid envelopes which will also be provided.

REIMBURSEMENT OF COORDINATOR AND ASSISTING PROCTORS

Each institution coordinator will receive an honorarium of $30.00 for his services in connection with the testing project. Following test administration, the coordinator should complete and sign the REIMBURSEMENT TO COORDINATOR form and return this form to Harvard using the reply envelope provided.

It is expected that at many institutions, where the total number of students to be tested does not exceed about 20, the coordinator will be singly responsible for test administration. At larger institutions, the coordinator may engage proctors (who may be

Appendix D-4 (continued)

may be either faculty members or graduate students) to assist him; one proctor may be engaged for each 20 students (or part of 20) over and above the first 20. Thus, for 21-40 students, one proctor is allowed; for 41-60 students, two proctors may be engaged. Each proctor will receive $10.00, and may receive this sum by completing and signing one of the "Reimbursement to Assisting Proctor" sections of the reimbursement form.

ADDITIONAL REMARKS:

After the test order form has been received for your institution, you will be sent a notice giving an INSTITUTION CODE NUMBER for the institution and also a specific range of STUDENT CODE NUMBERS to be assigned to the students at the time of testing. It is extremely important that all test materials--including the student questionnaire and the Modern Language Aptitude Test sheets--be marked with the correct code numbers. Your attention to coding details at the time of test administration will help to eliminate processing errors.

Questions which you may have in regard to the project may be addressed to:

College Foreign Language Testing Project
Graduate School of Education
Harvard University
Cambridge, Massachusetts 02138

If the situation appears to warrant a telephone call, please do not hesitate to call us collect at: Area Code 617, UN 8-7600, Ext. 3459.

We would like to take this opportunity to thank you for your assistance in the project and express our best wishes for a successful testing session.

Appendix D-5

568

ORDER FORM
*FOR ADMINISTRATION OF MLA FOREIGN LANGUAGE PROFICIENCY TESTS FOR TEACHERS AND ADVANCED STUDENTS
UNDER THE COLLEGE FOREIGN LANGUAGE TESTING PROJECT

IMPORTANT

This form must be returned to:
College Foreign Language Testing Project
Graduate School of Education
Harvard University
Cambridge, Massachusetts 02138
NO LATER THAN January 25, 1965

FOR OFFICE USE ONLY

Rec'd __________
Form __________
S.M. # __________
Inst. # __________
S.O. Rel'd __________
Code Notice Sent __________

PLEASE PRINT

Date of order__________ Test Administration Date__________
Institution__________ City, State__________
Ship to: Name and Title of Coordinator:__________
Institution__________
Street Address__________
City, State, and ZIP Code__________
REA Express Address (if different from above)__________

COMPLETE BATTERY contains Listening Comprehension*, Speaking*, Reading, Writing, Applied Linguistics, Civilization and Culture, and Professional Preparation

[FOR SENIORS IN TEACHER CERTIFICATION PROGRAMS ONLY]

Language:	FRENCH	GERMAN	ITALIAN	RUSSIAN	SPANISH
Number Ordered:	____	____	____	____	____

(Please order only the exact number required)

PARTIAL BATTERY contains Listening Comprehension*, Speaking*, Reading, and Writing

[FOR SENIORS NOT IN TEACHER CERTIFICATION PROGRAMS]

Language:	FRENCH	GERMAN	ITALIAN	RUSSIAN	SPANISH
Number Ordered:	____	____	____	____	____

(Please order only the exact number required)

TOTALS

Language:	FRENCH	GERMAN	ITALIAN	RUSSIAN	SPANISH
Totals:	____	____	____	____	____

(total number of examinees to be tested in each language)

*Please note that, for the Listening Comprehension and Speaking tests, ETS supplies Master Tapes which contain the stimulus materials. These are available on a 600-foot reel of tape, recorded at 3 3/4 i.p.s., or on a 1200-foot reel recorded at 7 1/2 i.p.s. Be sure to check the type desired if either of these tests is used.

MASTER TAPES for Listening and Speaking Tests (check type desired) ☐ 3 3/4 i.p.s. 600' reel ☐ 7 1/2 i.p.s. 1200' reel

*For the Speaking Test, five types of widely used blank tapes (or magnetic discs) are available for the recording of student responses. Be sure to check the type which is used with your recording equipment. If you have any doubts as to which of these is appropriate for your recording equipment, you should include with the test Order Form a description (including model number) of the type of recording equipment to be used.

BLANK TAPES for student responses to Speaking test (check type desired)
☐ 3 3/4 i.p.s. 600' reel ☐ 7 1/2 i.p.s. 1200' reel
☐ tape magazine (2 spindles) ☐ cartridge (1 spindle)
☐ magnetic discs (2 per student)

Please complete and sign "Terms of Test Use" on reverse side.

OVER

Appendix D-5 (continued)

TERMS OF TEST USE

In consideration of the fulfillment by Educational Testing Service of the order stated on the reverse side, the undersigned institution (hereinafter called "Test User") agrees as follows:

Test User will take full responsibility for administering the tests and for maintaining the security of the test materials. It is understood that the tests will be shown only to the students to whom they are administered.

Educational Testing Service shall ship all test materials via REA Express prepaid and the Test User shall return all test materials via REA Express collect, unless advance permission is obtained from the College Foreign Language Testing Project, Harvard University, to have these materials shipped to the institution and/or returned to Educational Testing Service by another type of carrier.

All tests and test materials, both used and unused, remain the property of the Modern Language Association of America and Educational Testing Service. In no event will Test User retain them for more than one week after the scheduled test administration date. The used answer sheets and test books, the unused test books, and the accessory materials are to be returned to Educational Testing Service immediately after the test administration. Student questionnaires administered under this program are to be returned direct to Harvard University (not to ETS) using the prepaid mailing envelopes provided.

It is understood that test score reporting under this project will be undertaken only by Harvard University (not Educational Testing Service) and will be restricted to: 1) a mailing of individual score reports to participating students in the early fall of 1965; and 2) a mailing at about the same time, of a roster of score reports to the chairmen of participating language departments, giving the test scores of students in their departments. Beyond these two services, neither Educational Testing Service nor Harvard University will undertake to furnish or honor requests to furnish at any time transcripts of test scores to students, institution officials, or other persons or organizations.

TEST USER

____
Name of Institution

By____
Authorized Signature of Coordinator

____
Title

____
Date

Appendix D-6

HARVARD UNIVERSITY
GRADUATE SCHOOL OF EDUCATION

COLLEGE FOREIGN LANGUAGE
TESTING PROJECT

APPIAN WAY
CAMBRIDGE, MASSACHUSETTS 02138

January 15, 1965

A LETTER TO COLLEGE SENIORS GRADUATING IN THE SPRING OF 1965 WITH DEGREES IN FRENCH, ITALIAN, SPANISH, GERMAN, OR RUSSIAN

We are sending you this letter (necessarily a form letter) to bring to your attention the fact that your school has been asked to participate in a project of foreign language testing to be conducted this spring. Under this project, sponsored by the U.S. Office of Education and administered here at Harvard, seniors in selected colleges and universities throughout the United States are being asked to take a series of four skill tests -- Reading, Writing, Listening, and Speaking -- in the language in which they are majoring (French, Spanish, Italian, German, or Russian). In addition, participating students are to be given a short form of a modern language aptitude test and also a questionnaire dealing with their language background, prior language courses taken, interests in foreign language study, and so forth.

For students who are certified by their institution as completing teacher preparation programs, three additional subtests are given in areas of knowledge particularly relevant to teaching: Civilization and Culture, Applied Linguistics, and Professional Preparation.

Results of the testing program will provide information about the general level of language proficiency of foreign language students throughout the United States near the end of their undergraduate training. Research of this type is also of great importance in the planning of more effective language teaching sequences at all levels of instruction.

Our purpose in writing this letter is to urge you strongly to participate in this study, that is, to take the series of four skill tests in the language in which you are majoring, the language aptitude test and background questionnaire, and, if you are following a teacher-preparation program, the three teaching-oriented tests mentioned above.

We will not attempt to minimize the task which we are requesting of you: participation in the program will involve giving approximately four hours* of your time on a spring testing date to be selected by your school; we realize that this is a considerable contribution on your part.

We would, however, like to emphasize the benefits to you which will be involved: First, test results, national norms, and other interpretative materials will be mailed to you as soon as available (at least by the early fall of 1965). With your scores and other accompanying information, you will be able to assess your own ability in each language skill by reference to the achievement of a nationwide sample of language majors. This opportunity to appraise your own strengths (and possibly, weaknesses) in particular areas of foreign language proficiency may be

---

*An additional two hours would be required for students taking the teaching-oriented tests.

Appendix D-6 (continued)

of importance to you in making plans for the use of the language in a vocation, in graduate study, or for other purposes. Furthermore, you will be able to report the results of these tests to graduate schools, business organizations, or others who may request them.

Second, we would like to point out that the tests are to be given free of charge; if you were to take the same tests "on your own," you would be required to pay the regular administration fee of $10.00 to $13.50, depending on the number of tests taken.

Third, achievement tests of the type to be administered are used at present and will be increasingly used by graduate schools, business concerns, and other organizations as part of their application and placement procedures. The opportunity to take tests of this type should be of considerable value to you in acquainting you with their general format and content.

A detailed description of the language tests, together with sample test items, will be found in the accompanying booklet, "A Description of the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students." Specific questions on the administration of the tests, the method of score reporting, and other details are answered at the end of this letter.

What we would like to request of you at the present time is that you make a commitment to participate in the testing program at your school on the date which will be announced in the near future. It is of importance to the success of the program that as many students as possible agree to participate, and we sincerely hope that you and your fellow language majors will all be willing to help us in this undertaking.

Sincerely yours,

John B Carroll

John B. Carroll
Project Director
Professor of Educational Psychology

Appendix D-6 (continued)

ADDITIONAL PROJECT DETAILS

WHO IS ELIGIBLE TO TAKE THE TESTS IN THIS PROJECT?

All seniors who are scheduled to graduate this spring from a participating institution and who are majoring in one of the languages in which the tests are given (French, Spanish, Italian, German, Russian) are eligible to take the four skill tests: Reading, Writing, Listening, Speaking. Seniors who are certified as completing a teacher preparation program are eligible to take in addition the three teacher-oriented tests: Civilization and Culture, Applied Linguistics, and Professional Preparation.

IS THERE A CHARGE FOR THE TESTS?

There is no charge either to the student or to the participating institution in this special testing project.

ARE THE TESTS DIFFICULT?

The tests have been deliberately designed to test language achievement at a rather high level of proficiency, such as would be expected of a person who has spent several years in language study. Thus, undergraduate foreign language majors, students starting graduate study in foreign languages, and beginning language teachers would find the tests appropriate. It is expected that individual students will find certain tests or parts of these tests quite easy for them, while other parts will prove more challenging. One real advantage of this series of tests is the fact that all of the important areas of language achievement are involved, allowing the student to determine those areas in which additional training or practice may be suggested, as well as those areas in which he is proficient.

HOW WILL TEST SCORES BE REPORTED?

Early in the fall of 1965, or possibly sooner, a report of scores for all tests taken by the student will be mailed to the student at the address specified by him at the time of testing. Along with the scores themselves a leaflet with tables of national test norms and explanatory information will be provided; these materials will help the student interpret his scores in each language area.

WHAT OTHER USES WILL BE MADE OF THE TEST SCORES?

In the fall of 1965, rosters will be sent to the chairmen of participating language departments, giving the test scores of students in their departments. It should be noted, however, that student performance on the tests will in no way affect college records.

The primary purpose for sending a score roster to the institution is to allow the foreign language staff to appraise the overall success of their language programs, rather than to examine the performance of an individual student as such.

Beyond the two services described above, individual test scores will not be released to students or institutions. It will not be possible to honor requests to furnish additional transcripts of test scores either to the student or to other persons or organizations. The student is free, of course, to report his own scores, if he so desires, to graduate schools, business organizations, and so forth, using the official form which will be supplied by this project.

OVER, PLEASE

Appendix D-6 (continued)

MAY CERTAIN TESTS BE OMITTED BY THE STUDENT?

All students who agree to participate in the testing project will be expected to take each of the tests in the series -- either four or seven tests depending on whether or not the student is enrolled in a teacher preparation program. Sufficient testing time will be scheduled by the institution, and ample rest periods will be provided.

WHERE MAY ADDITIONAL INFORMATION BE SECURED?

The foreign language chairmen at your institution have full information about the testing project. A testing coordinator has also been appointed by the institution; it is the responsibility of the coordinator to supervise the actual administration of the tests. The project staff at the Harvard Graduate School of Education will also be glad to receive questions or comments, which may be referred to:

College Foreign Language Testing Project
Graduate School of Education
Harvard University
Cambridge, Massachusetts 02138

Appendix D-7

FJ-251
Form 11 (revised)
2-65

College Foreign Language Testing Project

DIRECTIONS FOR ADMINISTERING THE QUESTIONNAIRE FOR FOREIGN LANGUAGE MAJORS

(It is suggested that you read over the questionnaire prior to the actual administration, in order to be better prepared to answer any questions that may arise.)

1) When the students are seated and ready to begin, make sure that they all have ordinary lead pencils (with erasers). Have some extra regular lead pencils on hand to lend to the students who may have forgotten to bring their own.

2) Distribute one questionnaire to each student.

3) Say: "Note that this is a four-page questionnaire.

"This questionnaire is a very important part of the project. You may find it somewhat lengthy, particularly if you have taken a large number of foreign language courses. We have scheduled a full hour during which you may work on this questionnaire, so there will be ample time for you to answer each item carefully.

"In answering the questionnaire, you may make use of the list you have brought to help you remember the language courses which you have taken. If you finish before the hour is up, bring the completed questionnaire to me at the front of the room. [Students may be allowed to leave the testing room quietly as they finish.]

"Directions are printed at the beginning of each part of the questionnaire. Please read all directions and answer every item carefully.

"Are there any questions? [Answer questions.]

"If you have any later questions, raise your hand and we will help you.

"Remember to answer all four pages. Please try your best to answer every item.

"You may begin answering the questionnaire."

[As the questionnaires are handed in, check them to see that they are complete and particularly that the name, address and correct institution and student code numbers are present. Please also check that all four pages have been completed.]

4) NOTES TO THE COORDINATOR:

In answering pages two and three, some students may have difficulty in determining where one "course" ends and another begins. A good working rule in this connection is that a new "course" is involved:

a) when there is a change of teacher, or
b) whenever there is a definite change in the form of instruction and/or material concerned.

In the rare event that a student has not finished the questionnaire in an hour, allow him extra time, if at all possible, on the same day. Be sure to include any late questionnaires in the regular shipment back to Harvard.

It is possible that some students will have taken more than 25 courses in their major language, and will require more working space to list and describe the extra courses. In this case the students may add 26, 27, etc. below the 25 [on page three] or they may use a separate blank sheet of paper.

RETURN ALL QUESTIONNAIRES TO HARVARD (NOT TO ETS) USING THE PREPAID MAILING ENVELOPE PROVIDED.

Appendix E-1

HARVARD UNIVERSITY
GRADUATE SCHOOL OF EDUCATION

COLLEGE FOREIGN LANGUAGE
TESTING PROJECT

APPIAN WAY
CAMBRIDGE, MASSACHUSETTS 02138

October 5, 1965

Memorandum to: The Foreign Language Majors who Took the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students in the Spring of 1965

On the bottom line of the above sticker you will find your Converted Scores on the Modern Language Association Foreign Language Proficiency Tests for Teachers and Advanced Students that you took in the spring of 1965. The scores are given in the following order: Listening Comprehension, Speaking, Reading, Writing, Applied Linguistics, Civilization and Culture, Professional Preparation. The top line of the sticker gives the code number for the institution, your examinee number, the language in which you were tested, and your name.

If no score or other indication appears for a given subtest, our records show that you did not take that subtest. If the symbols AAA, BBB, or CCC appear, there was an irregularity in either the administration or the performance of the test such that no valid score could be reported. These irregularities occurred particularly in the case of the Speaking test because of partial or total failure of the tape recording.

The enclosed leaflet will help you in the interpretation of your scores, all of which are given in the form of "converted scores" suitable for use with the norms in the leaflet. I wish to explain these norms to you. During the past several years, the Federal Government has supported the holding of so-called NDEA (National Defense Education Act) Institutes for teachers of foreign languages, generally during the summer months. The students in these Institutes are for the most part teachers of foreign languages in public secondary schools who have applied for admission to the Institutes and been accepted. Their travel and living expenses at the Institutes are provided by the government. Their previous training in foreign languages and experience in teaching varies widely; some are relatively recent college graduates, while others are teachers of long experience. The course of study at the Institutes includes refresher training in the language, and lectures in applied linguistics, civilization and culture and teaching methodology.

The students in the NDEA Institutes were tested shortly after arrival ("pretest") and again, with an alternate form of the test, at the termination of their study ("posttest"). As the enclosed leaflet explains, norms are given for both pretests and posttests. By following the directions on the front page of the norms leaflet, you can find the standings of your scores with respect to both the pretest and the posttest norms.

In view of the fact that these MLA proficiency tests (form "A") were administered to you under special testing arrangements made for this project only, it will not be possible for either this office or Educational Testing Service to honor future requests for score reports. You are free, of course, to report your own scores to various schools or other organizations. Students applying for state teacher certification should contact their own state education departments to determine departmental policy on the use of proficiency test scores for certification purposes.

Your scores on the Modern Language Aptitude Test, which most of you also took last spring, will not be furnished to you since these were collected only for statistical control purposes.

I wish to thank you, on behalf of the U. S. Office of Education, the Modern Language Association and Harvard University, for your cooperation in this study, and I hope that the results forwarded herewith will be of interest and use to you.

Sincerely,

John B. Carroll
Director, College Foreign Language
Testing Project

Appendix E-2

HARVARD UNIVERSITY
GRADUATE SCHOOL OF EDUCATION

COLLEGE FOREIGN LANGUAGE
TESTING PROJECT

APPIAN WAY
CAMBRIDGE, MASSACHUSETTS 02138

October 12, 1965

Dear Sir:

Enclosed are sheets containing the scores of students who were majoring in your department in the spring of 1965 and who took the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students at that time. Also enclosed is a leaflet giving norms for these tests as based upon the performance of foreign language teachers attending NDEA Institutes over the past several years.

As yet, the data from this testing have not been fully analyzed. Eventually, a full report on this testing will reach you. In the meantime, we feel that the norms that had been established for the teachers who attended NDEA Institutes provide a reasonably satisfactory basis for interpreting the performance of the FL majors that were tested at your institution, since it is of interest to compare the performance of the latter with that of teachers of foreign languages in the public schools. We do not have exact information on the nature of the sample of teachers that attended NDEA Institutes, but there is reason to believe that it is a fairly representative sample of all U. S. teachers of foreign languages in public secondary schools.

The teachers in the NDEA Institutes were tested shortly after arrival ("pretest") and again, with an alternate form of the test, at the termination of their study ("posttest"). As the enclosed leaflet explains, norms are given for both pretests and posttests. It will probably be of interest to you to compare your students' scores both with the pretest and with the posttest norms.

The identifying information at the top of the rosters gives, first, the three-digit code for your institution, followed by the language designation. Information for each student is then listed in the following order: student code; student name; test scores for Listening, Speaking, Reading, Writing, Applied Linguistics, Civilization and Culture, and Professional Preparation. The scores given are all "Converted Scores" appropriate for use with the accompanying norms leaflet.

If no score or other indication appears for a given subtest, our records show that the student did not take that subtest. The symbols AAA, BBB, and CCC are to be interpreted as follows:

AAA - Invalid Student Speaking Tape
BBB - Blank Student Speaking Tape
CCC - Irregularity reported by supervisor (this code is used only if it is found that the reported irregularity has so affected the student's performance on the test that his score cannot be considered valid).

These scores are also being furnished directly to the students concerned by mail to the addresses given by them. It is therefore not necessary for you to take any further action in reporting scores to your students.

I wish to thank you, on behalf of the U. S. Office of Education, the Modern Language Association, and Harvard University, for your cooperation in this study, and I hope that the results forwarded herewith will be of interest and use.

Sincerely yours,

John B. Carroll

John B. Carroll
Director

JBC:jc
Encl.

Appendix E- 3

MLA FOREIGN LANGUAGE PROFICIENCY TESTS FOR TEACHERS AND ADVANCED STUDENTS

INTERPRETATIVE LEAFLET AND TEST SCORES

	Pretest Scores							Posttest Scores						
	L	S	R	W	AL	C&C	PP	L	S	R	W	AL	C&C	PP
NDEA Inst. No.														

The Interpretative Leaflet has been prepared to help you interpret the scores you obtained on the MLA Foreign Language Proficiency Tests.

1. First, you will want to compare your scores for each pretest and corresponding posttest. The scores are expressed as converted scores so that pretest and posttest scores on corresponding tests are directly comparable.
2. Next, you will want to compare your scores with the scores of other institute participants by determining their percentile ranks from the norms tables.

The percentile rank for each score shows the percentage of institute participants who achieved scores below that score. For example, the percentile rank of a pretest score of 52 obtained on the French Listening Comprehension Test is 90. This means that a score of 52 on this test is higher than the scores obtained in pretesting by 90 per cent of the French institute participants. For another example, the percentile rank of a posttest score of 66 obtained on the Russian Writing Test is 75. This means that a score of 66 on this test is higher than the scores obtained in posttesting by 75 per cent of the Russian institute participants.

Not all percentile ranks and not all scores are actually shown in the tables. When the percentile rank for a particular score is not shown, you may obtain the percentile rank through interpolation. For example, a pretest score of 45 obtained on the Spanish Reading Test has a percentile rank of 65, and a pretest score of 47 has a percentile rank of 70. Since 46 falls one half of the way between the scores listed, the percentile rank should be considered to fall one half of the way between the given percentile ranks. In other words, the percentile rank of a pretest score of 46 will be 67.5. For another example, a posttest score of 105 obtained on the German Speaking Test has a percentile rank of 85, and a posttest score of 111 has a percentile rank of 90. Since 107 falls one third of the way between the given percentile ranks, the percentile rank of this score obtained in posttesting will be 86.7.

It can be anticipated that a few institute participants will achieve scores that have the same percentile rank on both pretest and corresponding posttest. About half of the remainder will achieve scores that have higher percentile ranks for the posttests. It is perfectly possible to have a gain in converted score from pretest to posttest, and still have a percentile rank on the posttest that is lower than the percentile rank on the pretest. This is due to the fact that the vast majority of institute participants showed a gain in converted scores on posttests. However, the likelihood of showing a gain in percentile rank from pretest to posttest depends in part upon the pretest score. In general, participants whose pretest scores fell below the fiftieth percentile are likely to have scores that have higher percentile ranks on the corresponding posttests.

The percentile tables for the French, German, Russian, and Spanish tests are based on the test results of all participants studying those languages in the 1961, 1962, and 1963 institutes. The percentile tables for the Italian tests are based on the test results of participants studying Italian in the 1961 and 1962 institutes; these percentile tables are labeled "tentative" because they are based on so few cases. The percentile tables for the Professional Preparation Test are based on the results of all participants in the 1961, 1962, and 1963 institutes.

Copyright 1964 by Educational Testing Service*
Princeton, New Jersey   Berkeley, California
All rights reserved

*Reprinted by permission of the Modern Language Association

Appendix E- 3 (continued)

MLA FOREIGN LANGUAGE PROFICIENCY TESTS for TEACHERS and ADVANCED STUDENTS

FRENCH

PRE-TEST

Percentile Ranks	Converted Scores				Appl. Ling.	Civ. and Culture
	Listening	Speaking	Reading	Writing		
99	56	119	67	72	65	67
97	55	111	65	68	62	62
95	54	106	63	66	59	59
90	52	98	59	61	56	55
85	50	93	56	58	54	53
80	48	89	54	55	52	51
75	46	85	51	53	50	49
70	43	83	49	50	49	47
65	41	80	47	48	48	46
60	39	77	45	46	47	45
55	38	74	44	44	46	44
50	36	71	42	42	44	43
45	35	68	40	40	43	42
40	34	66	39	38	42	41
35	33	63	37	36	41	40
30	32	61	36	35	40	39
25	31	58	35	33	39	38
20	30	55	33	31	38	36
15	29	51	32	29	36	35
10	28	47	30	26	35	33
5	26	43	28	24	32	31
3	25	40	26	22	31	30
1	22	36	24	20	29	27
No. Cases	4720	4494	4722	4723	4641	4640
Mean	38.1	72.1	43.3	43.0	44.9	43.6
Standard Deviation	8.8	19.3	10.8	12.9	8.1	8.5

POST-TEST

Percentile Ranks	Converted Scores				Appl. Ling.	Civ. and Culture
	Listening	Speaking	Reading	Writing		
99	58	122	67	72	72	70
97	57	115	65	68	68	67
95	56	110	64	66	66	65
90	54	104	61	63	63	61
85	53	100	58	60	61	59
80	52	97	56	58	59	57
75	50	93	54	55	58	55
70	49	90	52	54	56	54
65	48	88	50	52	55	52
60	46	86	48	49	53	51
55	44	84	46	47	52	50
50	43	82	45	46	51	49
45	41	79	43	44	50	48
40	40	76	42	42	48	46
35	38	74	40	40	47	45
30	37	71	39	38	46	44
25	35	69	37	36	45	43
20	34	65	36	34	44	42
15	33	62	34	31	42	40
10	31	57	32	29	40	38
5	30	51	30	25	36	36
3	29	48	29	23	35	34
1	26	42	26	20	32	31
No. Cases	4770	4577	4770	4769	4612	4611
Mean	42.8	81.0	45.7	45.7	51.1	49.3
Standard Deviation	8.6	17.8	10.5	12.6	8.9	8.7

GERMAN

PRE-TEST

Percentile Ranks	Converted Scores				Appl. Ling.	Civ. and Culture
	Listening	Speaking	Reading	Writing		
99	56	119	69	77	69	71
97	55	113	68	75	65	66
95	54	109	67	73	63	64
90	52	103	63	69	59	61
85	50	98	60	66	57	58
80	48	94	56	63	55	55
75	46	91	53	59	53	53
70	44	88	50	57	51	52
65	42	86	48	54	50	50
60	40	84	46	51	49	49
55	38	82	44	48	48	47
50	36	79	42	45	46	45
45	35	77	41	42	45	44
40	34	76	39	39	44	43
35	32	74	38	37	42	42
30	31	71	36	34	41	41
25		69	35	31	40	39
20	30	65	33	29	38	38
15	29	62	32	26	37	36
10	28	56	30	24	35	34
5	26	49	28	21	33	32
3	25	45	27	20	31	30
1	23	36	24	19	29	28
No. Cases	1102	1086	1103	1103	1102	1101
Mean	38.4	79.5	44.4	45.8	46.8	46.7
Standard Deviation	9.1	17.7	11.9	16.5	9.1	10.0

POST-TEST

Percentile Ranks	Converted Scores				Appl. Ling.	Civ. and Culture
	Listening	Speaking	Reading	Writing		
99	58	131	69	75	71	75
97	57	126	68	73	69	72
95	56	119	67	72	68	70
90	54	111	64	69	65	66
85	52	105	62	65	63	64
80	51	101	59	62	61	62
75	50	98	57	60	60	60
70	48	94	55	57	58	59
65	47	92	52	55	57	57
60	46	90	51	53	56	55
55	44	87	49	51	55	54
50	43	85	47	48	53	53
45	41	82	45	46	52	51
40	39	80	43	44	51	50
35	37	78	42	42	49	49
30	36	76	40	40	48	48
25	35	72	39	37	46	47
20	33	70	37	35	45	45
15	31	66	36	32	43	44
10	30	61	34	28	41	42
5	28	54	32	24	38	38
3	26	48		21	37	36
1	23	41	30	19	34	33
No. Cases	1256	1232	1256	1258	1179	1178
Mean	42.1	85.3	48.1	48.3	53.0	53.4
Standard Deviation	9.1	19.4	10.9	14.6	9.0	9.5

Appendix E- 3 (continued)

MLA FOREIGN LANGUAGE PROFICIENCY TESTS for TEACHERS and ADVANCED STUDENTS

RUSSIAN

	PRE-TEST							POST-TEST					
Percentile Ranks	Converted Scores				Appl. Ling.	Civ. and Culture	Percentile Ranks	Converted Scores				Appl. Ling.	Civ. and Culture
	Listening	Speaking	Reading	Writing				Listening	Speaking	Reading	Writing		
99	55	122	68	78	59	69	99	57	122	68	79	62	72
97	53	111	63	76	56	67	97	56	120	65	76	60	69
95	52	104	60	74	55	64	95	55	117	63	75	59	67
90	49	96	52	71	52	60	90	54	109	55	72	57	63
85	46	92	45	68	50	57	85	53	106	50	70	55	61
80	45	87	42	66	49	55	80	51	102	47	68	53	59
75	43	84	39	63	48	54	75	50	99	43	66	52	58
70	42	81	37	60	46	52	70	49	96	41	64	51	56
65	41	79	36	57	45	51	65	47	94	39	62		55
60	39	76	35	55	44	49	60	46	92	37	59	50	54
55	38	74	34	52	43	48	55	44	89	36	57	49	53
50	37	72	33	49	42	47	50	43	87	35	55	48	51
45	36	70	32	46	41	45	45	41	84	34	53	46	50
40		67	31	44	40	44	40	40	83	33	50		49
35	35	65	30	40		43	35	39	81	32	47	45	48
30	34	62	29	37	39	41	30	37	78		45	44	47
25	33	61	28	33	38	40	25		76	31	42	43	46
20	32	58	27	31	37	39	20	36	72	30	37	41	45
15	31	55	26	29	36	38	15	35	69	29	35	40	44
10	30	52	25	26	34	36	10	34	64	27	32	38	42
5	28	47	23	23	32	33	5	33	58	25	28	36	39
3	27	43	22	21	31	31	3	32	55	24	27	35	37
1	24	35	20	20	28	29	1	29	49	23	25	34	34
No. Cases	597	584	556	555	515	556	No. Cases	587	570	545	545	505	545
Mean	38.4	73.0	35.5	48.6	42.8	47.2	Mean	43.3	87.1	38.2	53.4	47.6	51.9
Standard Deviation	7.0	17.7	10.7	16.5	6.8	9.2	Standard Deviation	7.5	17.0	10.8	14.8	6.7	8.4

SPANISH

	PRE-TEST							POST-TEST					
Percentile Ranks	Converted Scores				Appl. Ling.	Civ. and Culture	Percentile Ranks	Converted Scores				Appl. Ling.	Civ. and Culture
	Listening	Speaking	Reading	Writing				Listening	Speaking	Reading	Writing		
99	55	118	66	74	64	72	99	55	116	66	76	68	75
97	54	111	63	72	59	69	97	54	111	63	73	66	73
95	53	108	61	69	57	67	95	53	108	62	72	64	71
90	51	101	57	66	54	63	90	52	101	58	68	61	68
85	49	94	54	62	52	60	85	51	97	56	65	59	66
80	48	90	51	59	50	58	80	50	93	53	62	57	64
75	46	86	49	57	48	56	75	48	91	52	60	56	63
70	45	83	47	54	47	55	70	47	88	50	58	54	61
65	43	80	45	52	46	53	65	46	85	48	56	53	60
60	41	77	44	50	45	52	60	45	82	47	54	51	58
55	40	74	42	48	44	50	55	43	80	45	52	50	57
50	38	71	41	46	43	49	50	42	78	43	50	49	56
45	37	69	39	43	42	48	45	41	76	42	48	48	55
40	36	66	38	42	40	46	40	39	74	41	46	47	54
35	35	63	37	40	39	45	35	38	71	40	44	46	53
30	34	60	35	38		44	30	36	69	38	42	44	51
25	33	56	34	35	38	42	25	35	66	37	40	43	50
20	32	53	33	33	37	41	20	34	63	36	38	42	48
15	31	49	31	31	35	39	15	33	59	34	36	40	47
10	29	45	30	28	34	37	10	31	55	33	33	38	45
5	28	40	28	25	32	35	5	30	49	31	30	36	42
3	27	38	27	23	31	33	3	29	45	29	28	35	40
1	25	33	25	21	29	30	1	27	39	27	25	33	37
No. Cases	4394	4271	4446	4444	4393	4342	No. Cases	4400	4316	4544	4547	4494	4443
Mean	39.3	72.1	42.2	46.2	43.3	49.6	Mean	41.8	78.2	44.6	50.2	49.4	56.2
Standard Deviation	8.0	20.4	10.2	13.7	7.8	9.7	Standard Deviation	7.6	17.5	9.5	12.8	8.5	8.9

Appendix E- 3 (continued)

MLA FOREIGN LANGUAGE PROFICIENCY TESTS for TEACHERS and ADVANCED STUDENTS

ITALIAN (tentative)

	PRE-TEST							POST-TEST					
	Converted Scores							Converted Scores					
Percentile Ranks	Listening	Speaking	Reading	Writing	Appl. Ling.	Civ. and Culture	Percentile Ranks	Listening	Speaking	Reading	Writing	Appl. Ling.	Civ. and Culture
99	52	116	62	75	60	73	99	51	126	64	80	66	76
97		115	61	73	59	67	97	49	123		78	64	71
95	50	112	59	72	57	62	95	48	121	63	77	63	67
90	48	109	58	71	55	60	90		114		74	61	65
85	47	103	57	69	53	58	85	47	113	60	72	60	64
80	45	99	56	67	51	56	80	46	111	59	71	58	62
75	44	98	54	64	50	54	75	45	110	57	69	56	61
70	43	96	53	60		53	70	44	107	54	66	55	60
65		94	52	59	48	52	65		105	52	62	54	58
60	42	93	50	56		51	60	43	104	50	61		57
55	41	92	49	53	47		55	42	102		60	53	56
50	40	90	47	52		50	50		100	47	57	51	54
45	39	89	45	50	46	49	45	40	98	46	52	50	53
40	38	87	41	46	45	48	40	39	97	45	51	48	52
35		85	39	43	44	47	35	38	94	44	49	46	51
30	37	82	38	41	43	46	30	37	91	41	47		50
25	36		36	38	41	44	25	36	88	40	42	44	49
20	35	78	33	37		43	20	35	86	38	38	43	47
15	34	70		33	39	42	15	33	79	34	37	42	46
10	32	65	32	31	37	39	10	32	75	33	35	41	43
5	30	61	29	29	35	37	5	30	72	29	33	37	40
3	28	55	27	27	33	35	3	28	68	28	30	36	38
1	26	46	25	24	29	33	1	27	65	25	26	34	36
No. Cases	64	62	64	64	64	64	No. Cases	64	62	64	64	64	64
Mean	40.0	88.2	45.2	51.1	46.1	49.8	Mean	40.5	98.0	47.5	55.3	50.7	54.5
Standard Deviation	6.0	15.4	10.3	14.4	6.5	8.0	Standard Deviation	5.9	14.8	10.7	14.7	7.8	8.5

PROFESSIONAL PREPARATION
ALL LANGUAGES

	PRE-TEST		POST-TEST
Percentile Ranks	Converted Scores	Percentile Ranks	Converted Scores
99	76	99	79
97	74	97	77
95	72	95	76
90	70	90	75
85	68	85	74
80	67	80	72
75	65	75	
70	64	70	71
65	63	65	70
60	62	60	69
55	61	55	
50	60	50	68
45	59	45	67
40	58	40	66
35	57	35	65
30	55	30	64
25	54	25	63
20	52	20	62
15	51	15	60
10	49	10	58
5	46	5	55
3	44	3	53
1	39	1	49
No. Cases	10771	No. Cases	10909
Mean	59.5	Mean	67.0
Standard Deviation	8.2	Standard Deviation	6.5

--Reprinted by permission of the Modern Language Association